Development Of Image Analysis Methods Applied To Collagen Imaged With Different Techniques

Thesis submitted in accordance with the
requirements of the University of Liverpool
for the degree of Doctor in Philosophy

by

Marion Alexandra Leibl

Department of Physics
University of Liverpool
June 2018

The work presented in this thesis is the author's except where other sources are named. No portion of the author's work described here has been submitted in support of an application for another degree or qualification in this, or any other, institute of learning.

Abstract

Development Of Image Analysis Methods Applied To Collagen Imaged With Different Techniques

This thesis is concerned with the development of analysis strategies for microscopic images of collagen, with the specific aim to minimise as far as possible the need for user input, thereby reducing potential bias. Images of collagen were chosen for analysis, as collagen is an easily available and biologically important molecule, whose organisation in tissues is of great interest due to its pivotal role in biomechanics, and its implication in many disease states. Quantifying its structure and organisation can facilitate better understanding of the physiological and mechanical role of collagen, as well as provide the basis for a more detailed knowledge of the structural foundation of the body.

We had access to a limited number of collagen images from porcine sclera obtained by Atomic Force Microscopy, Polarised Light Microscopy, Second Harmonic Generation Microscopy and Transmission Electron Microscopy. The challenges in developing analysis strategies arose from collagen itself, from the intrinsic differences in the data gathered by different microscopies, and from the presence of artifacts.

We developed successful methods to analyse data relevant to collagen characterisation from Atomic Force Microscopy, Polarised Light Microscopy and Second Harmonic Generation Microscopy images, and propose a quantification method where our analysis attempts were not successful.

This thesis aims to contribute to a growing body of work in image analysis, which can be applied across disciplines wherever image data are used. It is hoped that further development and refinement of the work presented here may be the focus of future work.

Acknowledgements

This project would not have been possible without the help, friendship, kindness and loyalty of many people, to all of whom I am deeply grateful. I would like to list every single one of them, but the list would be too long. This limitation does not diminish their importance.

First and foremost, I would like to thank my supervisors, Dr David Martin and Dr Steve Barrett, who have shown me unwavering loyalty thoughout these sometimes difficult years, and whose steady confidence in me has both puzzled and challenged me. I am grateful to them both for their professional and personal support, and especially to Dr Steve Barrett for many hours of conversations about images and other things.

Furthermore, I owe thanks to Dr Riaz Akhtar from the Biomechanical Engineering Group at the University of Liverpool for freely giving me access to many of his TEM and AFM images. Without these there would have been no project. My heartfelt thanks also go to Dr Chris Thrasivoulou, who generously made space for me to use his busy SHG microscope at UCL, and freely gave me his time for discussions and for practical and emotional support.

Tony Brandwood, Sean Williams, Val Tilston, Adam Bertram, Julie Haigh and Marion Pope from the Institute of Veterinary Science at the University of Liverpool made time in their pressured schedules to process my samples and were always happy to talk and advise me on any practical issues. I am indebted to them for their kindness.

I. A. Roberts of West Kirby, butchers extraordinaire, were most generous in providing me with fresh pig eyes and took great care in their initial dissection. I am very grateful to everyone in this friendly shop.

I would also like to thank Dr Mike Thomas and Dr Kirsty McKay for their interest in my adventures and their continued support over the years. Our discussions and their helpful suggestions on all aspects I have valued greatly, and their company when not talking about science has been a pleasure.

Most of all, I would like to say the biggest of thank-yous to Andy Newsam, who has been helpful in such a plethora of ways that listing them all would fill pages, and the list would still not be complete. From fundamental discussions to debugging my code, from believing in my abilities to being there with unwavering practical and emotional support when everything seemed to collapse, without him I am certain that this thesis would not have come anywhere near completion. Because of him, these years have been as productive, enjoyable, fulfilled and happy as an intrinsically stressful period of time in my life could possibly have been. For that, and for everything else, I dedicate this thesis to Andy.

To Andy

Contents

1 Introduction **29**
- 1.1 Considerations About Interdisciplinarity . . . 29
- 1.2 Computerised Image Analysis . . . 32
  - 1.2.1 Development of Digital Image Generation and Analysis . . . 32
  - 1.2.2 Advantages of Computerised Image Analysis . . . 35
- 1.3 This Project . . . 37

2 Collagen **41**
- 2.1 Collagen – A very brief introduction . . . 41
  - 2.1.1 Why is collagen important? . . . 41
  - 2.1.2 What makes a collagen? . . . 42
  - 2.1.3 Molecular Collagen . . . 44
  - 2.1.4 Collagen fibrils . . . 45
  - 2.1.5 Collagen fibres . . . 47
  - 2.1.6 Dynamic function . . . 48
- 2.2 Collagen in the sclera . . . 51
  - 2.2.1 Anatomy of the eye and sclera . . . 51
  - 2.2.2 Collagen types and matrix . . . 53
  - 2.2.3 Scleral disease . . . 54

3 Atomic Force Microscopy **57**
- 3.1 Atomic Force Microscopy . . . 57
  - 3.1.1 Principles . . . 57
  - 3.1.2 The Atomic Force Microscope . . . 57
  - 3.1.3 Imaging Modes and data interpretation . . . 58
- 3.2 AFM Images and Analysis . . . 61
  - 3.2.1 AFM Images . . . 61
  - 3.2.2 Biological Significance . . . 62
  - 3.2.3 Analysis strategy . . . 62
- 3.3 Macro Analysis . . . 67
  - 3.3.1 Macro outline . . . 67
  - 3.3.2 Procedure *DrawRedDot* . . . 70
  - 3.3.3 Procedure *DrawGreenDot* . . . 70
  - 3.3.4 Functions *degtorad* and *radtodeg* . . . 71
  - 3.3.5 Procedure *FindStep* . . . 71
  - 3.3.6 Procedure *AngledSobel* . . . 71
  - 3.3.7 Procedures *DrawROIquad1, -2, -3, -4* . . . 72
  - 3.3.8 Procedure *GetPeriod* . . . 73
  - 3.3.9 Procedure *DrawMainROI* . . . 75
  - 3.3.10 Procedure *DrawAllROIs* . . . 75
  - 3.3.11 Procedure *AFMmeasure* . . . 76
  - 3.3.12 Macro *AFMmeasureWithLoop* . . . 77
  - 3.3.13 Analysis variables . . . 77
- 3.4 Evaluating the macro . . . 79

3.4.1 Frequency measurements . . . 79
3.4.2 Fibre diameter measurements . . . 80
3.5 Future work . . . 91
3.6 Conclusion . . . 92

4 Polarised Light Microscopy 93
4.1 Principles of Polarised Light Microscopy . . . 93
4.1.1 Polarised Light . . . 93
4.1.2 The Polarised Light Microscope . . . 95
4.1.3 Polarised Light Microscopy of Biological Materials . . . 98
4.2 PLM Images and Analysis . . . 98
4.2.1 Preparation Methods . . . 99
4.2.2 PLM Images . . . 99
4.3 Analysis Strategy . . . 101
4.4 Macro Development . . . 104
4.4.1 Metrics . . . 106
4.4.2 Macro outline . . . 110
4.4.3 Procedure *ACFtilespart* . . . 110
4.4.4 Procedure *ACFtileswROIsizes* . . . 112
4.4.5 Macro *ACFtileswROIsizesallimages* . . . 112
4.5 Macro Evaluation . . . 112
4.5.1 Impact of ROI size . . . 112
4.5.2 Testing Thresholds . . . 113
4.5.3 Analysis of sample images . . . 118
4.6 Future work . . . 126
4.7 Conclusion . . . 128

5 Second Harmonic Generation 131
5.1 Principles of Second Harmonic Generation Imaging . . . 131
5.1.1 Second Harmonic Generation . . . 131
5.1.2 The SHG Microscope . . . 134
5.1.3 SHG of Collagen . . . 135
5.2 SHG Images and Analysis Strategy . . . 136
5.2.1 SHG Samples . . . 136
5.2.2 SHG Images . . . 137
5.2.3 Thresholding . . . 138
5.2.4 Analysis Strategy . . . 141
5.3 Results Comparison . . . 142
5.4 Future work . . . 149
5.5 Conclusion . . . 150

6 Transmission Electron Microscopy 151
6.1 Principles of Transmission Electron Microscopy . . . 151
6.2 The Images . . . 152
6.2.1 Histograms of Entire Images . . . 157
6.2.2 Histograms of Segmented Images . . . 158
6.2.3 Line ROIs Across Boundaries . . . 161
6.2.4 Fourier Filtering . . . 165
6.2.5 Outlines by Eye . . . 171
6.2.6 Degrees of Image Degradation . . . 173
6.3 Quantification of Success - the Quality Evaluation Method (QEM) . . . 177
6.4 Future work . . . 182

6.5 Conclusion . . . 183

7 Bonus Chapter **185**
7.1 Play with the PLM Analysis . . . 185
7.1.1 Millennium 2 Simulation . . . 185
7.1.2 River Deltas . . . 187
7.1.3 Trees . . . 188
7.1.4 Walls . . . 190
7.1.5 Breakfast Cereals . . . 191
7.1.6 Vegetables . . . 192
7.2 Play with the AFM Analysis . . . 193
7.2.1 Fence . . . 194
7.2.2 Roof . . . 194
7.2.3 Wall . . . 197

8 Conclusions **199**

A Brief Summary of Some Image Analysis tools **203**
A.1 Thresholding and Density Slicing . . . 203
A.2 Erosion and Dilation . . . 203
A.3 Kernel Filters and Convolution . . . 206
A.4 Sobel filter . . . 206
A.5 Watershed . . . 207
A.6 Particle Analysis . . . 208
A.7 Rotating Mask . . . 210
A.8 Hough Transform . . . 212
A.9 Fast Fourier Transform - FFT . . . 213

B Macro Codes **218**
B.1 AFM Macro . . . 218
B.2 PLM Macro . . . 231
B.3 TEM Macro . . . 237

References **241**

List of Figures

1.1 Illustration of the difficulty with obtaining reliable data through just looking. The shape above the horizon (A) seems of a uniform "grey" colour, whereas the shape below the horizon (B) seems "white". Now cover up the centre where the two areas A and B meet with your finger, and discover that A and B are in fact the same colour. (From [49]) 36

2.1 Illustration of the ubiquitous presence of collagen in human tissues. Major collagenous tissues are labelled, however, due to being an important constituent of all connective tissue, it pervades the whole body at small scales. 43

2.2 Processing of procollagen to collagen and assembly into fibril. Adapted from [60]. 46

2.3 Staggered arrangement of fibrillar collagen leads to a periodic banding pattern in a fibril. a) Schematic drawing of the gaps and overlaps giving rise to a 67 nm D-period. b) TEM image of the same banding in a collagen fibril. Adapted from [68]. 47

2.4 Schematic of the hierarchical assembly of collagen into tendon. Adapted from [58]. 49

2.5 Multimicrovacuolar collagenic dynamic absorbing system (MCDAS), the ubiquitous, fractal-like structure that is proposed to envelop and permeate the entire body, from video "Strolling under the skin" ([77]). a) Schematic representation of the MCDAS around a tendon, b) endoscopy image of an MCDAS structure 50

2.6 Anatomy of the human eye: only major landmarks are labelled. Adapted from [78]. 52

2.7 Schematic of cross-section through the posterior sclera. The dotted line separates sclera (above) from other structures (below) that are not part of the sclera but are drawn to provide context. Holes represent sections through blood vessels. 52

3.1 The principle of the Atomic Force Microscope. A laser beam detects the deflection of the cantilever as the tip scans over the sample surface. A feedback loop readjusts the tip-sample distance and an image of the surface is reconstructed. 58

3.2 Convolution of volumes, a common tip artifact. The AFM tip is less sharp than a surface feature and it is the tip geometry rather than the surface geometry that constrains the path. a) Tip scans across real surface (blue) and describes a path (red line); b) The reconstructed image shows mirroring of the tip geometry. The slopes are artifacts of the tip geometry. 59

3.3 Typical force-distance curve in Peakforce QNM. On the curve the parts used for the different default channels of Peak force error, DMT modulus, adhesion, dissipation and deformation are labelled accordingly. Modified from [90] 61

3.4 Overview of height (a), and peak force error (PFE) (b) channels of the images 2E.010, 2A.010, 2PA.034, 3A.009 and 3EB.035 used for algorithm development. All images are 256×256 pixels. 63

3.5 An example AFM image with its 8 channels. An identical line ROI has been drawn in each channel to show fibre edges. a) height; b) peak force error; c) DMT modulus; d) Log DMT modulus; e) adhesion; f) deformation; g) dissipation; h) Sneddon modulus. The corresponding plot profiles are shown in figs. 3.6 and 3.7. 65

3.6 Plot profiles of channels 1 – 4 (a – d) in Fig.3.5. a) height; b) peak force error; c) DMT modulus; d) Log DMT modulus. 66

3.7 Plot profiles of channels 5 – 8 (e – h) in Fig.3.5. e) adhesion; f) deformation; g) dissipation; h) Sneddon modulus. 66

3.8 Another example AFM image with its 8 channels. This time, an identical line ROI has been drawn in each channel to show banding along the fibre axis. a) height; b) peak force error; c) DMT modulus; d) Log DMT modulus; e) adhesion; f) deformation; g) dissipation; h) Sneddon modulus. . 67

3.9 Plot profiles of channels 1 – 4 (a – d) in Fig.3.8. a) height; b) peak force error; c) DMT modulus; d) Log DMT modulus. 68

3.10 Plot profiles of channels 5 – 8 (e – h) in Fig.3.8. e) adhesion; f) deformation; g) dissipation; h) Sneddon modulus. 68

3.11 Typical AFM image of collagen-rich tissue. The fibre orientations vary, the fibres run above and beneath each other and the boundaries between them fade in and out. a) original scan; b) lines drawn over visible edges by hand. Note the degree of fragmentation of the lines. 69

3.12 Brief outline of the macro algorithm. The names of the procedures at the relevant steps are written in blue. For clarity, some procedures have been omitted from the overview. 69

3.13 Illustration of how edges are emphasised and de-emphasised by the angle-constrained Sobel filter, depending on the measured angle of the user ROI. 72

3.14 Comparison between the original image (a), the conventional (b) and the angle-constrained Sobel filter (c). Note the absence of banding in c. 72

3.15 Correspondence of angles with quadrants in a unit circle. 73

3.16 Illustration of how the end points of the secondary ROIs are derived from the user ROI's angle a. 74

3.17 Examples of the extended ROI's plot profile and the plot profile of the one-dimensional FFT of this. The grey circle indicates the mirror point, where the drawn line ROI's plot profile ends and is extended by its mirror image. The main peak in the FFT at around 0.076/pixel corresponds to a length scale of about 13 pixels, which is not easily recognised in the line profile of the ROI. The smaller peak at around 0.155 corresponds to about 7 pixels and is likely a harmonic of the first. The obvious periodicity of around 20 pixels in the line ROI profile corresponds to the small peak just under 0.05/pixel in the FFT plot. 75

3.18 Sequence in which the macro draws the secondary ROIs. The black arrow represents the user ROI, the grey lines the secondary ROIs and numbers indicate the sequence. 76

3.19 The output of the AFM measurements macro: a) original image with the user ROI drawn in; b) angle-constrained Sobel filter applied to image shown in a); c) copy of original image with green dots indicating found fibre edges and red dots to show where the ROIs have been drawn; d) text file with Fourier frequencies; e) text file with half diameters, diameters, average and standard deviation; f) plot profile of line ROI extended to next 2^n; g) frequency plot of data from d). 77

3.20 After the angle-constrained Sobel filter has been applied, a larger, more pronounced peak identifies the fibre edge (b), compared to the unfiltered image (a). 78

3.21 Finding the right value for *steplength* is crucial for maximising the grey value difference detected from a given peak. The step length shown in green detects the maximum grey value difference, whereas the blue and purple step lengths are too small and too large, respectively, to find substantial differences. 79

3.22 These two examples show how a wide ROI can positively or adversely affect how clearly periodicity shows in a plot profile. In 1a) the ROI is drawn perpendicular to the banding, which is clearly visible in the plot profile. In 1b) the same ROI is widened to 30 pixels. The banding is even more obvious. In 2a) the ROI is drawn at an angle across the same fibre, which changes the plot profile somewhat, but still shows some periodicity. In 2b) the same ROI as in 2a) is widened to 30 pixels, and the periodicity in the plot profile is lost. 80

3.23 Some examples of ROI plot profiles and corresponding FFT plot profiles from macro output. The ROI plot profiles are in the left column, the corresponding FFT plot profiles in the right column. The frequency corresponding to collagen banding is highlighted in red. The slight shift away from this frequency in the last example may be due to the ROI being drawn at a slight angle to the fibre axis. 81

3.24 Histogram of *stepthresh* grey values resulting in DifAvg values below 1.5 with DifN=6. Data from one image (2PA-034-2). n=180. 82

3.25 Histogram of *steplength* pixel values resulting in DifAvg values below 1.5 with DifN=6. Data from one image (2PA-034-2). n=180. 83

3.26 Histogram of *ROIwidth* pixel values resulting in DifAvg values below 1.5 with DifN=6. Data from one image (2PA-034-2). n=180. 83

3.27 Success rate of the macro over all parameter sets. Column of N=0: the number of parameter sets that found 6 edges and thus returned no zero; column of N=1: the number of parameter sets that found 5 edges and returned 1 zero; and so on until column N=6: number of parameter sets that found no edges and returned 6 zeros. The weighting is strongly towards low N numbers, indicating a good success rate. Only measurements with DifAvg ¡ 1.5 pixels are included. 84

3.28 Histogram of the DifAvg values, a) from 0 to 100 pixels. Bin size is 1 pixel. b) from 0 to 2 pixels. Bin size is 0.2 pixels. n=4977. Note the significant peak within the first five bins in b), corresponding to a < 1 pixel difference between manual and macro measurements. 85

3.29 Histogram of *stepthresh* values resulting in DifAvg values below 1.5 with DifN=6. All images and all values for *steplength* and *ROIwidth* are included in the data. n=706. 86

3.30 Histogram of *steplength* values resulting in DifAvg values below 1.5 with DifN=6. All images and all values for *stepthresh* and *ROIwidth* are included in the data. n=706. 86
3.31 Histogram of *ROIwidth* values resulting in DifAvg values below 1.5 with DifN=6. All images and all values for *stepthresh* and *steplength* are included in the data. n=706. 87
3.32 At the most successful step lengths, 25, 30 and 35 pixels, the *steplengths* producing DifAvg values below 1.5 appear to be relatively uniformly represented. 87
3.33 At the most successful step lengths, 25, 30 and 35 pixels, the *ROIwidths* producing DifAvg values below 1.5 also appear to be relatively uniformly represented. 88
3.34 A ROI gradually elongated along the same fibre (fibre 1). (Read left to right and top to bottom.) The ROI origin is kept fixed. As the ROI centre and end positions move parallel to edge features as they appear and disappear, the algorithm succeeds and fails in finding the correct edge. (See green dots.) In this case, the fibre edges show sufficient contrast and most attempts by the macro succeed. 68 edge points were identified correctly, 3 were identified incorrectly, and 1 was not found. 89
3.35 A ROI gradually elongated along the same fibre (fibre 2). (Read left to right and top to bottom.) The ROI origin is kept fixed. As the ROI centre and end positions move parallel to edge features as they appear and disappear, the algorithm succeeds and fails in finding the correct edge. (See green dots.) In this second case, the fibre edges often are too faint and cannot be identified by the macro. 62 edge points were identified correctly, 6 were identified incorrectly, and 4 were not found. 90
3.36 Difference between the average diameter measured by the macro and the manually measured diameter, plotted for each elongation step. At each step, the environment of the points at which the secondary ROIs are drawn alters and leads to varying success in edge finding. The average deviation from manual measurement was 7.9 pixels in the case of fibre 1, and 9.6 pixe;s in the case of fibre 2. 91

4.1 Light represented simplified as an electromagnetic wave, in which the electric field vector oscillates at a certain frequency and amplitude in one plane, and the magnetic field vector equivalently in a perpendicular plane. (Adapted from [101].) 93
4.2 The principal polarisation types. The light is assumed to emerge at right angle from the page in the direction of the viewer. In the case of linear polarisation (a), the e-vector telescopes back and forth in one direction. The ratio of the compound vectors remains constant. In the case of elliptical polarisation (b), the e-vector describes an ellipse. The ratios of the compound vectors change with the phase. In the case of circular polarisation (c), the e-vector describes a circle. The ratios of the compound vectors change with the phase, but the vector sum always adds up to a constant value. 96
4.3 Circularly polarised light. In red: Path described by the sum e-vector; in blue: path described by the vertical component of the e-vector; in green: path described by the horizontal component of the e-vector. This is a special case of circularly polarised light, in which the phase shift between the two component vectors is $\pi/2$. (Adapted from [102].) 96
4.4 Schematic drawing of the light path through a simple polarised light microscope. (Adapted from [104].) 97

4.5 Example of section through a porcine eye ball stained with a standard histological stain (H&E). The areas where images were taken were the cornea (C), limbus (L), medial of the eye (M), posterior eye (P), lateral to the lamina cribrosa (LC) and optic nerve head (ON). 100
4.6 Example of PLM image at 10x lens magnification. Red staining is from collagen birefringence enhanced by picrosirius red. The scale bar is shown here for information and is the same for all images examined, but will not be shown in images used for analysis. 100
4.7 Visualisation of the ACF: (a) Two copies of an image are superposed, one copy is shifted in the x- and y-direction by Δx and Δy; A = area of overlap of circles; (b) the correlation between the image and its copy is measured as $A(x')$, the area of overlap as a function of x- displacement. Taken from [114], with permission. 102
4.8 Simple shapes and their ACFs in grey scale and colour look-up table (LUT). A black on white ellipse as starting image in a) has an ACF whose contour lines are also ellipses, see b). This is more obvious when grey values are rendered in colour (c). In contrast, an irregular black on white image (d) has an only slightly anisotropic ACF, but does not replicate the starting shape contours (e), again shown more clearly in colour in f). 103
4.9 Example of the difference between the ACF of a real and a simplified image of similar shapes. a) real image; b) real image ACF; c) plot profile along line in b; d) simplified image; e) simplified image ACF; and f) plot profile along line in e. Note the difference in low contrast texture between the ACF images. 104
4.10 Example of how the presence of different shapes affects the shape of the ACF and its thresholded particle. Column 1,2 and 3 show different constellations of simplified shapes. Row a: the black and white shapes; row b: the ACFs of row a; row c: ACFs from b thresholded at a grey value of 83. The shape of the particle in row c, and how many are visible, depends on the chosen threshold value. See fig. 4.11. 105
4.11 Dependence of the shape of the central particle of an ACF on the chosen threshold. The ACF 3b from fig. 4.10 was thresholded at the grey values given in the top left corners. There is strong dependence of particle shape on the threshold value. 105
4.12 Output data from the Particle Analysis: area (black) is the sum of all black pixels in the central particle, angle (grey) measured counterclockwise and relative to the horizontal, major and minor axis (red and blue, respectively) are the longest diameter and its perpendicular drawn through the centre. . 106
4.13 Example sections of the compound images of thresholded ACF particles. Section of original image (a) and ACF images of same section at ROI size 16 pixels (b), 32 pixels (c), 64 pixels (d), 128 pixels (e) and 256 pixels (f). . 107
4.14 Calculation principle of the sum vector's length and angle. a) Angles α, β and γ are measured angles, already multiplied by 2. a,b and c are the corresponding vectors. b) Vector addition results in a sum vector with length s and angle δ. The *Average Angle* is calculated as $\delta/2$. *AngleR* is calculated as $s/3$, with 3 vectors added. 109
4.15 Brief outline of the macro for PLM analysis. The user input is minimal with decisions only about the range of ROI sizes. Although this macro has a small number of separate procedures, all the main parts listed here are carried out by procedure *ACFtilespart*. Auxilliary procedures are not mentioned here. 111

4.16 The large ROI (a) shows little axial anisotropy in its thresholded ACF (b), however a smaller ROI (inset in a) from the centre of the image shows a much more anisotropic central particle. Threshold in both cases is gv = 74. 112

4.17 Example of sampling artifacts from a linear dummy image. A manually created dummy image with parallel, but not completely equidistant, lines at 30 degrees (a) and ACF particles at ROI size 16 (b), ROI size 32 (c) and ROI size 64 (d). Note the variation in the thickness, angles and areas, greatest at ROI 16, but still present to a lesser extent in ROI 32. 114

4.18 Explanation of the diversity shown in fig. 4.17: two ROIs of 16 pixels (1 and 2) were taken from different locations of the dummy image and show the variation of image shapes. The original and thresholded ACFs result in substantially different measurements for Area, Angle and Major and Minor Axes. 114

4.19 Average Angle measurements of the Linedummy image for ROI sizes 16 to 128. The image lines are angled at 30 degrees. Although the angles of the ACF particles vary at every ROI size due to ROIs covering more than one line, they do not deviate widely from 30 degree. 115

4.20 The two images that were used for testing thresholds. G-LC2 has a denser fibre structure than E-L2. 115

4.21 Comparison of the influence of threshold values on *AxanisAve* values. Thresholds shown from grey values 100 to 200. 116

4.22 Comparison of the influence of threshold values on *AxanisDev* values. Thresholds shown from grey values 100 to 200. 116

4.23 Comparison of the influence of threshold values on *AngleR* values. Thresholds shown from grey values 100 to 200. 117

4.24 For clarity, in 4.21, 4.22 and 4.23 the thresholds shown range from grey value 100 to 200, although the thresholds tried range from 60 to 240. The threshold graphs outside the 100 - 200 range showed erratic behaviour and were therefore omitted in those figures. Here the threshold graphs excluded from the above graphs are shown for the example of AngleR. Each point in the graphs consist of three data points per image. 117

4.25 Histogram of KS probabilities for comparison of all image pairs. Each image was compared to the others at the same ROI size and the KS probabilities for each ROI size were averaged for this image pair. This averaged KS probability was entered in the histogram, a) for *Axanis*; b) for *AngleDev*. 406 data points in each case. 119

4.26 Typical examples of PLM images from group A, B and C, as classified by visual appearance. All image sizes are 1022 x 722 pixels, corresponding to 689 μm x 487 μm (A), 691 μm x 488 μm (B) and 522 μm x 390 μm (C). . . 121

4.27 *AxanisAve* for groups A, B and C. A: n = 3, B: n = 12, C: n = 8. 122

4.28 *AxanisDev* for groups A, B and C. A: n = 3, B: n = 12, C: n = 8. 122

4.29 *AngleR* for groups A, B and C. A: n = 3, B: n = 12, C: n = 8. 123

4.30 *AxanisAve*, *AxanisDev* and *AngleR* curves for group A. Note that *AxanisAve* stays constant throughout all ROI sizes, *AxanisDev* decreases only slightly, and *AngleR* rises from a medium value to almost 1 at large ROIs. The error bars are standard deviations from the variation between images, n = 3. 124

4.31 *AxanisAve*, *AxanisDev* and *AngleR* curves for group B. Here *AxanisAve* is at first similar to A, but slightly rises at larger ROI sizes, *AxanisDev* is similar to A, and *AngleR* starts at a higher value than in A, but likewise rises to almost 1 at large ROIs. The error bars are standard deviations from the variation between images, n = 12. 124
4.32 *AxanisAve*, *AxanisDev* and *AngleR* curves for group C. The *AxanisAve* behaves very similar to A, and the *AxanisDev* curve follows the *AxanisAve* trend almost exactly. The *AngleR* begins at a lower value than in A and B and rises linearly to a value of only 0.8 at large ROIs. The error bars are standard deviations from the variation between images, n = 8. 125
4.33 Unclassified image GM2. 126
4.34 *AxanisAve*, *AxanisDev* and *AngleR* from unclassified image GM2. 126
4.35 *AxanisAve*, *AxanisDev* and *AngleR* from unclassified image GM2, compared to group C. Note that the corresponding graphs for *AxanisAve* (blue and green) follow similar trends but are further apart than one standard deviation (error bar). The situation is similar for *AxanisDev* (red and brown). In the case of *AngleR* (yellow and light blue), however, the yellow graph for GM2 lies mostly within one standard deviation from the light blue curve of group C. 127

5.1 Schematic representation of the generation of second harmonic photons. Excitation light hits the target material, which converts two excitation photons of wavelength λ into one emission photon of wavelength exactly $\lambda/2$. 132
5.2 Phasematching of the SHG signal with the excitation beam (red wave): a) the source of the second harmonic generation in a sample thinner than the wavelength of light, Δk is close to zero, forward propagated waves are in phase with the excitation waves; b) second harmonics generated in a thicker sample (on the scale of the wavelength of light and thicker), Δk is larger than zero, the backward propagated waves do not have well-matching phases and therefore destructive interference can destroy some of the signal. Red dots represent harmonophores. (Adapted from [126].) 133
5.3 Schematic representation of an SHG microscope. (PMT = photomultiplier tube, GLP = Glan laser polariser) (Adapted from [129].) 135
5.4 Example of SHG image at 10x lens magnification. The scalebar is shown here for information, and the scale is the same for all SHG images. 137
5.5 Typical examples of SHG images from group A, B and C, as classified by visual appearance. These images are the SHG equivalents of the PLM images shown in fig. 4.26. It can be seen that, while the SHG images show more detailed structure at small length scales, there is considerable similarity to the PLM images. Image size is 1022 x 722 pixels, corresponding to 739 μm x 522 μm 139
5.6 PLM (a) and SHG (b) images of similar eye regions and both in visual group B. In the SHG image smaller fibres often make up the larger ones, whereas in PLM the larger fibres are a single block. 140
5.7 Threshold range for analysis of SHG images. The values at grey value 130 and at grey values greater than 160 meet the exclusion criteria of an *Angle* of 0, and *Axanis* of 1 (round) and a pixel area of less than 10, and would therefore be excluded from analysis of the macro data. It can be seen that the thresholds that yield meaningful data range from about 40 to 120. A threshold of 80 was chosen for the analysis. 140

5.8 Schematic drawing of the theoretical relationship between different imaging modalities and the underlying structure. The basis of all data are the real biological structures present in living things. Only some of these structures will be markers for tissues, conditions, age stages or any other groupings. Any imaging modality will only be able to capture some aspects of the biological reality. Choosing the right modality will ensure that some of the data captured contain information on structures relevant for classification into groups, and therefore of interest for analysis (areas a, a1 and a2). Most likely, the imaging modality will also capture information that is not relevant for classification and analysis (areas b, b1 and b2). Each imaging modality, by nature of its different physical principles and tissue preparation requirements, will capture some of the same and some additional structures of the underlying biology. A robust and versatile analysis strategy will be most successful when the data to be analysed fall into area a, where the structure both yields information relevant for classification, and can be captured by more than one imaging modality. (Artifacts are not taken into account in this schematic, which would correspond to areas where the imaging modality circles reach into areas outside the biological structure.) . 142

5.9 *AxanisAve* for SHG groups A, B and C. A: n = 3, B: n = 10, C: n = 9. . . 143

5.10 *AxanisDev* for SHG groups A, B and C. A: n = 3, B: n = 10, C: n = 9. . . 143

5.11 *AngleR* for SHG groups A, B and C. A: n = 3, B: n = 10, C: n = 9. 144

5.12 *AxanisAve* for PLM groups A, B and C. A: n = 3, B: n = 12, C: n = 8. . . 144

5.13 *AxanisDev* for PLM groups A, B and C. A: n = 3, B: n = 12, C: n = 8. . . 145

5.14 *AngleR* for PLM groups A, B and C. A: n = 3, B: n = 12, C: n = 8. 145

5.15 *AxanisAve*, *AxanisDev* and *AngleR* curves for SHG group A. Note that *AxanisAve* stays constant throughout all ROI sizes, *AxanisDev* decreases only slightly, and *AngleR* rises from a medium value to almost 1 at large ROIs. The error bars are standard deviations from the variation between images, n = 3. 146

5.16 *AxanisAve*, *AxanisDev* and *AngleR* curves for SHG group B. Here *AxanisAve* is at first similar to A, but slightly rises at larger ROI sizes, *AxanisDev* is similar to A, and *AngleR* starts at a higher value than in A, but likewise rises to almost 1 at large ROIs. The error bars are standard deviations from the variation between images, n = 10. 146

5.17 *AxanisAve*, *AxanisDev* and *AngleR* curves for SHG group C. The *AxanisAve* behaves very similar to A, and the *AxanisDev* curve follows the *AxanisAve* trend almost exactly. The *AngleR* begins at a lower value than in A and B and rises linearly to a value of only 0.8 at large ROIs. The error bars are standard deviations from the variation between images, n = 9. . . 147

5.18 *AxanisAve*, *AxanisDev* and *AngleR* curves for PLM group A. Note that *AxanisAve* stays constant throughout all ROI sizes, *AxanisDev* decreases only slightly, and *AngleR* rises from a medium value to almost 1 at large ROIs. The error bars are standard deviations from the variation between images, n = 3. 147

5.19 *AxanisAve*, *AxanisDev* and *AngleR* curves for PLM group B. Here *AxanisAve* is at first similar to A, but slightly rises at larger ROI sizes, *AxanisDev* is similar to A, and *AngleR* starts at a higher value than in A, but likewise rises to almost 1 at large ROIs. The error bars are standard deviations from the variation between images, n = 12. 148

5.20 *AxanisAve*, *AxanisDev* and *AngleR* curves for PLM group C. The *AxanisAve* behaves very similar to A, and the *AxanisDev* curve follows the *AxanisAve* trend almost exactly. The *AngleR* begins at a lower value than in A and B and rises linearly to a value of only 0.8 at large ROIs. The error bars are standard deviations from the variation between images, n = 8. . . 148

6.1 Schematic drawing of a Transmission Electron Microscope. An electron beam is accelerated and focussed onto a biological specimen stained with heavy metals. The unscattered, transmitted electrons are visualised by a fluorescent screen or CCD camera below. (Adapted from [139].) 152
6.2 All TEM images for analysis of cross sections: Images were classed by quality: From 2L-04 to MC004 *Green* images (best), from 2E-05 to MLt003 *Yellow* images (somewhat worse), from 2E-08 to MLt005 *Orange* images (considerably worse), from 2L-03 to AC005 *Red* images (worst). The scales of these image differ, and due to the absence of calibration data in the file headers, the scales of the images could not be ascertained. 156
6.3 Image 2L-07 (green) and histogram of grey values. Note that the distribution of grey values is close to a Gaussian distribution with no additional peaks. 158
6.4 Image 2P-07 (green) and histogram of grey values. The distribution of grey values is still relatively close to a Gaussian distribution, but there is an additional expansion on the lighter slope. 159
6.5 Image 2L-03 (red) and histogram of grey values. Also here the distribution of grey values is close to a Gaussian distribution with no additional peaks. . 159
6.6 Image 2L-07: a) original image; b) manually drawn mask; c) image of blobs only with background removed; d) image of background only with blobs removed; e) histogram of blobs; f) histogram of background 160
6.7 Image 2P-07: a) original image; b) manually drawn mask; c) image of blobs only with background removed; d) image of background only with blobs removed; e) histogram of blobs; f) histogram of background 160
6.8 Image 2L-03: a) original image; b) manually drawn mask; c) image of blobs only with background removed; d) image of background only with blobs removed; e) histogram of blobs; f) histogram of background 161
6.9 Boundaries between background and blobs: a), b) and c) clear boundaries; d), e) and f) unclear boundaries; a) and d) from image 2L-07; b) and e) from image 2P-07; c) and f) from image 2L-03 162
6.10 Image 2L-07: line ROIs of various thicknesses (1, 5, 10, 15 and 20 pixels) drawn through the background-blob boundary. First column shows the line ROI and plot profile through what is considered a clear boundary. Second column shows the same drawn through what is considered an unclear boundary. 163
6.11 Image 2P-07: line ROIs of various thicknesses (1, 5, 10, 15 and 20 pixels) drawn through the background-blob boundary. First column shows the line ROI and plot profile through what is considered a clear boundary. Second column shows the same drawn through what is considered an unclear boundary. 163
6.12 Image 2L-03: line ROIs of various thicknesses (1, 5, 10, 15 and 20 pixels) drawn through the background-blob boundary. First column shows the line ROI and plot profile through what is considered a clear boundary. Second column shows the same drawn through what is considered an unclear boundary. 164

6.13 Determination of the length scale of the noise. a) ROI drawn inside blob with plot profile. b) ROI drawn in background with plot profile. The smallest noise in on the length scale of approximately 5 pixels. 166

6.14 An original image and its FFT image. There are no dark areas in the FFT image that might indicate a particularly prevalent noise frequency. 166

6.15 a) Low pass filter at a threshold of 5 pixels, and b) inverse FFT image of a), original without frequency components of 5 pixels and lower. c) High pass filter at a threshold of 5 pixels, and d) inverse FFT image of c), containing only frequency components up to 5 pixels. The inset in d) shows a part of d) at 6x magnification. 167

6.16 a) grey value histogram of the inverse FFT image of a low pass filter with a threshold at 5 pixels (elements larger than 5 pixels); b) grey value histogram of the inverse FFT image of a high pass filter with a threshold at 5 pixels (elements smaller than 5 pixels). The orange line indicates the peaks at identical locations on the grey value spectrum. 168

6.17 After a low pass filter at 5 pixel threshold was applied, line ROIs of various thicknesses (1, 5, 10, 15 and 20 pixels) were drawn through the same blobs as in fig. 6.10. Again, the first column shows the line ROI and plot profile over what is considered a clear boundary. Second column shows the same drawn over what is considered an unclear boundary. The plot profiles have been smoothed, but the boundaries are not more recognisable than in fig. 6.10 169

6.18 Inverse FFT images of low pass and high pass filters at various frequency thresholds. a) to c) low pass filters; d) to f) high pass filters; a) and b) threshold at 10 pixels; b) and d) threshold at 15 pixels; c) and f) threshold at 25 pixels 170

6.19 Sampling across a larger part of the image to illustrate differences between blobs and background. a) part of an inverse FFT image after a low pass filter at 5 pixel threshold, with black line indicating line ROI location; b) plot profile of line ROI in a). The areas where the line ROI crosses a blob are underlined in colour. Note the similarity in noise levels between blobs and background, and the non-obvious boundaries. 170

6.20 a) Example FFT image of one of our image set and b) its radial profile. The inset shows a diagrammatic representation of three image circles that, averaged, make up a radial profile plot point each. There is no peak in the radial distribution, suggesting that there is no particular, clustered noise frequency. 171

6.21 Image 2L-07: convincing blob outlines traced by hand. Most blobs have large stretches that can be traced, but few have no breaks in their outlines. Especially adjacent to other blobs lines appear to break up. 172

6.22 Image 2P-07: convincing blob outlines traced by hand. Here most blobs have only one gap in their outlines, but the gap does not seem to show a preference for always appearing close to neighbouring blobs. Few have an unbroken outline. 172

6.23 Image 2L-03: convincing blob outlines traced by hand. Very few clear outlines exist in this image. Most of the impression of blob shapes seems to be created by the visual system extrapolating to ovals. There are many places where an edge is possibly present, but could be in several locations. . 173

6.24 Dummy images that demonstrate progressive image degradation. The first three images contain only one aspect of degradation: noise (n), broken outline (bo), and low contrast (lc); the next three combine noise and a broken outline (n = bo) and low contrast and a broken outline (bo + lc) and noise and low contrast (n + lc). The last image combines all three, noise, broken outline and low contrast (n + bo + lc). The noise consists of single pixels at the two grey values of background and blob, unsmoothed. . 174
6.25 Hough Transform applied to images in fig. 6.24. For labels refer to fig. 6.24. 174
6.26 Gaussian smoothing filter applied to images in fig. 6.24. The kernel size is 5x5 pixels. For labels refer to fig. 6.24. 175
6.27 Gaussian-smoothed images from fig. 6.26 thresholded at grey value 128. For labels refer to fig. 6.24. 175
6.28 Rotating Mask applied to the image in fig. 6.24. Neighbourhood radius 15 pixels. For labels refer to fig. 6.24. 176
6.29 Rotating Mask image in fig. 6.28 thresholded at grey value 128. For labels refer to fig. 6.24. 176
6.30 Flow diagram of example analysis strategy that achieved partial success. Here the parameters are diameter of the Rotating Mask, threshold grey value, "count" number for erosion and dilation, and iterations for erosion and dilation, and the cut-off points for size in the Particle Analysis. 180
6.31 Two images for a simple illustration of the QEM. a) is a relatively good quality image (image 2L-04) and b) is more degraded (image 2L-03). The results for the images above, after the analysis routine outlined in fig. 6.30 has been applied, are shown in c) and d). In this particular case the diameter of the Rotating Mask was 31 pixels, threshold grey value was the mean grey value of the starting image, "count" number for erosion and dilation was 3 in both cases, and iterations for erosion and dilation was 10, also in both cases. The manually outlined blobs for both images are shown in e) and f). Because of the small area and for the purpose of this demonstration the particles touching the edges have been included. 181

7.1 Images of the Millennium 2 Simulation of the formation of structure in the universe at different red shifts, with their "analysis results". These images show a 2D slice of the original 3D data. The scale is 100 Mpc/h each side, the image size is 1024x1024 pixels. a) red shift z = 6.2; b) z = 2.07; c) z = 0.99; d) z = 0; larger red shifts denote a universe from an earlier time, thus z = 0 represents the present day 187
7.2 Satellite images of river deltas, with their "analysis results". Image sizes around 600 pixels. a) Hawooi And Kikori Deltas (Fly River) in Papua New Guinea (credit: NASA); b) Ganges in India and Bangladesh (credit: NASA); c) Lena in North-Eastern Siberia (credit: NASA) and d) Mississippi in Louisiana (credit: ESA) 188
7.3 Three images of trees (a - c) and part of a tree with fractal-like properties (d). Image sizes are on the scale of about 1000 pixels. Since a tree is a 3D object, some of its structure is obscured, and therefore the results will to a certain extent depend on the position of the photographer. 189
7.4 Images of an irregular stone (a) and a more regular brick wall (b). Image size for both 1280x960 pixels. 190
7.5 Three types of breakfast cereal: a) muesli; b) granola; c) jumbo oats. Image size 512x512 pixels. 192
7.6 Two kinds of vegetables with structure on different scales. a) broccoli; b) cauliflower. Image sizes 512x512 pixels. 193

7.7 Metal fence with regular spacing in front of wall with wooden planks at approximately right angles to the fence. 194

7.8 AFM analysis of a section of fig. 7.7. Note that the image has been turned into grey scale and then inverted, since the macro was developed to find dark edges rather than light. In this case the line ROI has been drawn horizontally, so that the "fibre width" result measures the width of the wooden planks, correctly at 35 pixels (not shown). a) image showing the location of the line ROI and the green dots drawn where an edge is detected (circled in black); b) plot profile along the line ROI; c) Fourier spectrum of the plot profile: the red peak at 0.022/pix corresponds to the largest repeat of fence grates at 45 pixels, the blue peak at 0.125/pix corresponds to the width of the metal bars at 8 pixels. The large peak at 0.0625/pix corresponds to 16 pixel and is likely a harmonic. 195

7.9 AFM analysis of a different section of fig. 7.7. This time the line ROI is drawn vertically, so that the "fibre width" becomes the spacing of the metal bars, which is again measured correctly at 45 pixels. a) image showing the location of the line ROI and the green dots (circled in black) where an edge is detected; b) plot profile along the line ROI; c) Fourier spectrum of the plot profile: There is a peak at 0.029/pix (second from left), which would correspond to the plank width of 35 pixels, but there are also other equally strong peaks, probably due to the many variations in grey value in the planks along the ROI. 196

7.10 Old roof tiles of a terraced house. Although the blur was not intentional, it may have helped by smoothing out irregularities that could interfere with correct "fibre width" measurement. a) results image showing the location of the line ROI and the detected edges; b) plot profile of the line ROI; c) Fourier spectrum of the plot profile: Since the line ROI only covers two tile lengths (each tile is approximately 175 pixels long, corresponding to about 0.006/pixel), this periodicity is not clear in the Fourier spectrum. 197

7.11 A brick wall (grey values inverted to facilitate edge finding). a) results image with line ROI location and found edges indicated by green dots (circled black for clarity), four of the six edge points coincide correctly with the brick edges; b) plot profile of the line ROI showing a brick length of 160 pixels (corresponding to a frequency of 0.006/pix); c) Fourier spectrum of the plot profile, the strong peak at 0.125/pix corresponds to a length of 8 pixels, about the width of the cement between the bricks. 198

A.1 Examples of thresholding and density slicing. a) original image (artifact, presumed to be houseplant pollen); b) histogram of a); c) image a) thresholded at grey value 74. All pixels with a higher grey value have been set to black, all pixels below 74 to white; d) image a) density sliced between grey values 110 and 150. The values between 110 and 150 are shown in red. These pixels can either be given a set grey value, leaving the remaining pixels unchanged, or can be left unchanged and the remaining pixels can be replaced by a given grey value. 204

A.2 Example of binary erosion. Only black pixels are considered in this operation. The numbers in the black pixels reqresent the number of white pixels touching. When the count is set to 3, a black pixel is turned white only if the number of white pixels touching it is at least 3. Therefore, the only pixel that remains black is the central one, whose neighbourhood only contained two white pixels. 205

A.3 Example of binary dilation. Only white pixels are considered in this operation. The numbers in the white pixels represent the number of black pixels in the immediate neighbourhood. When the count is set to 4, white pixels with at least 4 black pixels in the neighbourhood are turned black. In this case, only the central pixel whose neighbourhood contained 6 black pixels is turned black. 205

A.4 Example of the convolution of an image with a kernel (in this case a simple smoothing kernel). The grey values of the image pixels are indicated in pink. When the kernel filter is applied to this image (which is part of a larger image), the central pixel's new grey value becomes $(1x218 + 1x207 + 1x220 + 1x203 + 1x204)/5 = 210$. The grey values in the other pixels are also changed, since the larger source image was smoothed as a whole. 206

A.5 Kernels of the Sobel filter: the kernel emphasising vertical edges is on the left, the kernel emphasising horizontal edges on the right. 207

A.6 a) Original image containing many steep grey value gradients ; b) after application of the Sobel filter: only the edges remain. 207

A.7 Demonstration of the watershed segmentation method. An irregular blob is shown in a). For the distance map, b), the pixels inside the blob are assigned grey values as a function of their nearest distance to a boundary. The ultimate points in c) are derived from sequential erosion of the blob edges and are the local maxima of b). The watershed lines are shown in d), for clarity in e) the watershed lines are shown together with the ultimate points, and in f) the distance map, the ultimate points and the watershed lines are overlaid to show their relationship. 208

A.8 The Measurement Options dialogue box of the Particle Analysis and Measure functions. 209

A.9 The Particle Analysis Options dialogue box. 209

A.10 Example of an image after the Particle Analysis has been run, and the Results window with the measurements from that image. 210

A.11 Principle of a Rotating Mask. The environment of a pixel is divided into two parts, each of which is examined with respect to its homogeneity. As the subset of pixels covers a blob, the inhomogeneity increases. The subset that covers the blob to the greatest extent is likely to have the largest degree of inhomogeneity and its counterpart is likely most tangential to the blob. Thus, the new pixel grey value is assigned in relation to this counterpart, for example the average grey value. 211

A.12 Flow diagram of the Rotating Mask implementation in the TEM analysis. Standard deviation of grey values is used as inhomogeneity criterion, and the mean grey value of the complement of the ROI half with highest inhomogeneity is used for the pixel value in the filtered image. 211

A.13 Example of the effect of the Rotating Mask. a) original image with no sharp edges; b) after the Rotating Mask has been applied, some edge features can be discerned. 212

A.14 User interface of special HT analysis that detects size distribution in ImageSXM. Sizes from 15 to 25 pixels were chosen. 212

A.15 Illustration of the Hough Transform: a) dummy image with arbitrary black shapes on a white background, on average 40 pixels in diameter; b) Hough Transform result for 19 pixel diameter circles (one of a stack from 15 to 25 pixels diameter, not shown), note that a central dark accumulator forms for objects that are close to the size of the applied diameter, while for objects of a different size the accumulator does not form; c) cumulative result: circles drawn round the accumulator points with the diameter at which they were found; d) overlay image of original with the result of the search for circles from 15 to 25 pixels in size. 213
A.16 Various filters based on manipulation of the FFT image. a) unmodified FFT image of b), low frequencies are in the centre, high frequencies toward the periphery; b) the original image; c) low pass filter: a black circle is drawn around the centre of the FFT, this acts as a mask that tells the software only to use the data underneath the black area for the iFFT; d) iFFT of c), only low frequencies (large objects) remain in the image; e) high pass filter: the black mask covers the entire FFT image except the area of the low freqencies around the centre; f) iFFT of e), only small scale structure is left in the image; g) filter based on thresholding the FFT image: this allows frequencies to be selected that occur above a certain power, independent of their length scale; h) iFFT of g) 215
A.17 Examples of masks used to separate image features by frequency. a) unaltered FFT image of the original (a) in A.18); b) mask selecting the "ladder" at around 45 degrees; c) mask selecting the unbroken diagonal through the centre; d) mask excluding the diagonal through the centre; e) mask excluding the "ladder" at 45 degrees; f) mask selecting the set of two double points. (Note that the masks are drawn in outline only, to allow to see what would be underneath. 216
A.18 Images corresponding to the FFT images and masks in fig. A.17. a) original image, unaltered; b) compared to a), this image has lost all but the banding pattern in the collage fibres. hence the 45 degree "ladder" in the image above originates from the collagen banding; c) in comparison to a), here all features have been lost, except the overall shapes of the long fibres, hence the diagonal through the centre represents the large-scale fibre shapes; d) the overall fibre shapes have been removed from this image, including the light and dark areas at large scales, leaving mainly the banding in a uniform background; e) the fibre banding is the only thing that has been removed from this image, leaving the long fibre shapes and the rest of the image features intact; f) the double points in fig. A.17 are sharp and clearly defined, pointing at very exact frequencies. Isolating these reveals a wave throughout the image near 50 Hz, commensurate with electrical noise. . . . 217

List of Tables

1.1 Some examples of the widespread use of computerised image analysis. . . . 35

2.1 Classification of collagen types by structure and function. Adapted from [58]. 44

4.1 Brief overview of the metrics used in the analysis, *AngleR*, *AxanisAve* and *AxanisDev*. 110
4.2 Comparison of unclassified image GM2 with groups A, B and C for all metrics (*Simvals*) in units of standard deviations. 127

6.1 Automatic and manual measurements of numbers of blobs, area percent and axial anisotropy in the images shown in fig. 6.31. 179
6.2 QEN for images 2L-04 and 2L-03 when analysed by the routine outlined in fig. 6.30 using one parameter set. 179

Chapter 1

Introduction

1.1 Considerations About Interdisciplinarity

Image analysis, when applied to images from the life sciences, can be regarded as a branch of biophysics. The field of biophysics is very wide, as it broadly describes any area in which biological systems are described or investigated with the help of methods used in the physical sciences, or where the aim is to better understand their physical characteristics, as opposed to their chemical aspects. This diversity necessarily leads to a degree of vagueness, when talking about biophysics. Biophysics spans areas as far apart as the study of the atomic structure of macromolecules, the study of the behaviour of flocks of birds, examination of the mechanics of cell shapes, the investigation of human gait, to the use of computers to analyse images, with many more fields in between. This diversity shows two things: the usefulness to biology of input from physics, and the universal applicability of aspects of physics in living systems ([1]).

After the time when the Natural Sciences made no distinction between different areas of knowledge of the natural world, biology and physics have developed without much overlap or cross-fertilisation until recently, at least in Western science. For a while it was even thought that living and non-living matter obeyed a different set of natural laws. Discoveries of the chemical compounds in living things led to the development of the field of biochemistry, and since then it has been attempted to explain most of biology by chemical interactions alone. Soviet science appears to have had an interest in biophysics somewhat earlier than the West, but the Cold War and the language barrier prevented this from spreading. An interesting account of the history of electricity in biology can be found in "The Body Electric" by Robert O. Becker [2].

One of the most interesting areas for the cooperation between physics and biology is at the level of the cell. Although historically there have been repeated forays into attempts to explain biological phenomena at the cellular level through physical mechanisms (for example [3], [4], [5], [6]), this has not usually found widespread acceptance amongst the scientific mainstream. Some ideas that were proposed could not yet be properly tested because of experimental difficulties, or because appropriate equipment was not yet available. Living systems are typically smaller, wetter, more fragile, more complex, and further away from equilibrium, than the standard physics experiment can cope with. Likewise,

standard experimental setups in cell biology had been developed for chemical, but much less so for physical characteristics. This has started to change in recent decades. With the ability to miniaturise many pieces of equipment, the physical investigation of living things has become more feasible. Also, computational advances have allowed a degree of complexity, accuracy and refinement that would have been out of reach only twenty years ago. Opening these doors has led to a greater openness to the idea that physics may be able to contribute significantly to biological understanding, and also physicists have become increasingly interested in the complex challenges that biology presents.

Cultures and fashions develop in science as much as in any other area of human endeavour. In cell biological research, a strong culture towards the chemical aspects of life developed when the power of explaining interactions through chemical structures and reactions was realised. This has led to a proliferation of data about thousands of biomolecules and their variants, their production and processing and also their regulation and feedback mechanisms. While this is to some extent necessary and interesting information, it seems like a game of diminishing returns, in which the wood starts to disappear behind the trees. An overarching view of larger patterns is much needed and perhaps slowly starting to emerge, and the fresh outside view and mathematical tools of physics may be helpful in formulating a description of these patterns ([7]).

Even at the textbook level, the treatment of protons, electrons and other entities at or near the quantum scale could benefit from an outside perspective. In biology, such matter is usually regarded as particles with entirely classical behaviour. For instance, the ubiquitous charge distribution across a lipid membrane is often seen as no more than charges in different cellular compartments, but their function as a capacitor is often not appreciated. Such examples are numerous. Living beings are full of charged particles, flowing liquid and interfaces, they are subjected to and generate mechanical stresses that shape and influence their outer forms and locomotion, they generate light in many cases and interact with it in sophisticated ways ([8]), they form part of larger assemblies that follow specific rules and they are part of ecosystems, which they shape and are shaped by. Many of these aspects invite a physical description and mathematical treatment.

The reasons for why the description of living beings has neglected their physical aspects is as much in the culture as it is in the widespread lack of interdisciplinary education. An average university course in a biological subject does not include enough physics for the student to learn to think in a physical way. Likewise, most students in the life sciences do not get taught enough mathematics to develop their own creative ideas for its use.

An interdisciplinary approach would benefit any science, not only the life sciences. The current university education, with its emphasis on increased and earlier specialisation and relatively short degree courses at the expense of generality, represents an obstacle to the growth of any science. A solid grounding in all the natural sciences would allow researchers to think of solutions outside their own immediate field, even if they might need

specialists from other fields to implement these ideas. Communication is the foundation of science as a common human endeavour, and as the conventions of language and of thinking differ substantially between different fields, the difficulty of understanding and being understood by researchers in other fields sometimes presents a substantial barrier. A more interdisciplinary graduate education might go some way to alleviate this problem. Especially a common understanding of mathematical basics in the sciences would be of great benefit to all. A pertinent example of this may be the discussion around meaningfully demonstrating statistical significance by the use of p-values in the life sciences, which has been going on for several years and has not yet been resolved (see for example [9]). Furthermore, best practice could be shared more widely. Open publication of raw data, for example, which is standard practice in some areas of physics, could be an approach other research communities might also benefit from, although the current funding model in the biological sciences, heavily reliant on the pharmaceutical industry, might make this difficult to achieve.

It is generally thought that physicists can add understanding to other fields, but that the reverse is not true. However, it might be beneficial even for physics to delve into new areas. The trend towards simplifying reality in physics, evocatively illustrated by the image of a spherical cow, is contrasted by the image of a categorising biologist who spends her life carefully cataloguing hundreds of species of plants, perhaps, creating a zoo of details. Both are of course caricatures, but they do contain some grains of truth. Although simplification is a useful first step in understanding basic principles, real-life situations are rarely simple and more complex processes may wash out or at least modify the expected outcome. Perhaps simplification of intrinsically complex systems might overlook underlying, interesting properties. Experiencing the wet messiness of biological systems might help in learning to formulate descriptions that take into account more factors and find out how matter behaves when it is warm and wet.

Questions such as 'What is life?' have occupied people's minds for centuries. Many other questions have been merely intriguing, for example the question about the nature of health and disease, the navigation abilities of birds and other animals, the workings of the senses or the nature of consciousness. I believe that the contributions of physics in particular can progress our understanding of these phenomena more than any other science currently can. These contributions may on the one hand be conceptual, and on the other hand also practical, since advances in computing power in recent years, often spear-headed by physics projects, provide the opportunity now for the first time to carry out experiments in the life sciences that were hitherto out of reach.

An understanding of biophysics will, I believe, lead to some of the most exciting developments and insights in biology in the future. It is also possible that understanding the principles of living systems from a physical perspective might help physicists understand more fully the properties of matter so far only investigated in non-living systems.

1.2 Computerised Image Analysis

1.2.1 Development of Digital Image Generation and Analysis

Analysing images is something we do naturally. All senses intact, we predominantly rely on our visual system for making sense of the world around us. A large part of our brain (the visual cortex) is dedicated to analysing images. Like most bodily functions, this happens to a large degree out of our awareness. When photons excite action potentials in the cone and rod cells in the retina, this in itself has no meaning for us. It is only between the retina and the visual cortex that this information gets processed and integrated, and - together with other stimuli and previous experience - this allows us to attach meaning (recognition of objects, patterns, motion, etc.) to the raw data of the incident photon pattern. Not only the visual cortex, but also other brain systems are involved in this. For example, in a condition called "blind sight", the person affected is functionally blind and has no awareness of seeing. However, when faced with the choice between an image of something threatening and non-threatening, the person will always choose the non-threatening image, without being consciously aware why or of having actually seen the image [10]. This indicates that how the brain processes visual information is far more complex than mere pattern recognition.

We analyse images not only automatically, but also intentionally. Probably ever since we, as a species, have had an awareness of past and future, we have tried to use one to predict the other. We analyse cloud formations to predict the weather. We extract data concerning the whereabouts of other animals from patterns on the ground. We look at each others' faces and interpret the other person's mood, intention and thoughts. We look at cars coming down the road and analyse their likely trajectory and speed to gauge the risk to our lives when crossing the road. All these are complex analyses of visual data that we have learned to carry out more or less consciously. Analysing what we see has been crucial to our lives, whether millions of years ago or now.

Before the invention of photography, the visual experience of one person could not be shared easily with another person, unless the person happened to be there at the same time, looking at the same thing. Painting and drawing were often used to get around this problem and record events in order to share them with others who were in different places, or to preserve the information for posterity. In this process, the making of the image and the analysis are not clearly distinguished. In order to translate a three-dimensional situation or process into a two-dimensional, static representation, the artist has to constantly analyse and make decisions about what to include or leave out, and how to find a good representation of what he or she might see. Substantial analytical skill was needed for this, especially since the outcome was intended in many cases to provide the "raw data" for further analysis.

When photography was invented, it also translated a three-dimensional reality into a two-dimensional representation, but unified the set of limitations and compromises associ-

ated with it. Where every individual artist previously had to make constant decisions and analyse the world before them from his or her own idiosyncratic, subjective view point, photography always uses the same intrinsic "algorithm" for how the third dimension is reduced to two. Thus, in photography some information is also lost, but in a constant, reproducible way, and not many individual decisions are involved.

For an observer, there is no fundamental difference between looking at the moving, three-dimensional world around and looking at a two-dimensional representation of it. An important advantage of the representation, however, is its permanence. In representing reality, a reduced version of it can be stored and frozen in time, it can be shared with other people from exactly the same view point, and the experience of it can be repeated again and again. This opens the door to a degree of objectivity that was impossible before. Yet more objectivity was gained through the digitisation of images. In analog photography, depicted objects could be counted and measured by an observer, but extracting reliable numerical information from an image became much more feasible when the image was divided into discrete building blocks (pixels), each with its own discrete value (often numbers from 0 to 255 corresponding to grey value, or intensity of the red, green and blue channels). A digital image, therefore, is not only a picture to look at, but a matrix of numbers that can easily be manipulated with mathematical tools.

The possibility of analysing images with mathematical methods opened the door for computerised image analysis, as a set of algorithms that allow the manipulation, classification and numerical assessment of image content. There is a conceptual difference between image processing, in which images are manipulated in various ways and prepared for analysis, and image analysis proper, in which image content is analysed and data extracted. In practice, however, processing is usually such an integral part of analysis that the two can be regarded as a unit, as the kind of processing that is necessary is often specific to the analysis required.

Not only the nature of images changed with the progress of technological developments, but also the tools with which the images are acquired. The human eye can see electromagnetic radiation with wavelengths between around 380 nm to 750 nm, and has a resolving power in the region of a few hundredths of a degree ([11]). Furthermore, it can only see from the vantage point of where it is feasible for the human body to be. The invention of optical lenses was the first technological advance in this area, and extended the eye's resolving power significantly. In more recent times, optical instruments that are sensitive to non-visible parts of the electromagnetic spectrum were and continue to be developed, thus widening the parts of the natural world that can be "seen" substantially. Moreover, cameras can be mounted on submarine vessels, drones, aeroplanes and satellites, and even inserted into living organisms, thus unfolding many new vantage points from which images can be taken. All these developments have led to a proliferation of new kinds of data, which are rich in information waiting to be extracted and analysed.

In addition to these novel kinds of data, the exponential increase in computing power in the last decades has allowed the acquisition of data in general, and image data in specific, at a faster and faster rate. "Big Data" are continuously generated by a multitude of sources, whether through the internet, businesses, security, healthcare, the media or through scientific experiments like the Large Hadron Collider or the planned Large Synoptic Survey Telescope, as just a couple of examples. Especially automated processes can often gather data much faster than they can be analysed, or even stored, and without fast and efficient analysis algorithms to match this rate, information is either not optimally used or lost altogether.

Digital imaging has become ubiquitous, not only in research, but also in daily life. Although digital images are everywhere, computerised image analysis is still not nearly used as widely and in all areas where it would have potential. The possible reasons for this are discussed below. While much more scope exists for computerised image analysis in general, there are also fields where digital images are analysed routinely, or at least where there are well-established niches. To give an impression of the variety of fields in which digital image analysis is used, table 1.1 lists some areas of research and a few exemplary publications to illustrate the point.

Area	Topic	Reference
Life Sciences	Colocalisation of labels in microscopy images	[12]
Histology	Automated analysis of tissue sections	[13]
Medicine	Automated assessment of breast cancer cells	[14]
Marine biology	Measuring characteristics of sea grass leaves	[15]
Apiology	Estimating strength parameters of bee colonies	[16]
Plant Science	Estimating plant disease severity	[17]
Animal behaviour	Measuring colour change in rock pool gobies	[18] [19]
Microbiology	Quantification of biofilm removal	[20]
Astronomy	Analysis of star clusters in Hubble Space Telescope images	[21]
Military	Hyperspectral image data analysis	[22]
Oil industry	Estimating oil recovery in sandstone reservoirs	[23]
Geology	Study of laminated sediments	[24]
Meteorology	Website dedicated to meteorological monitoring	[25]
Environmental Science	Desertification monitoring	[26]
Environmental Science	Measuring plant cover in the Arctic	[27]
Environmental Science	Monitoring shoreline changes	[28]
Environmental Science	Measuring aerosol characteristics	[29]

Environmental Science	Analysis of sand grain sizes	[30]
Environmental Science	Dynamics of land use and hydrology	[31]
Environmental Science	Digital analysis of radar data	[32]
Security	Face recognition	[33]
Security	Iris recognition	[34] [35]
Security	Automated finger print identification	[36]
Archaeology	Quantitative monitoring of rock art sites	[37]
Palaeontology	Distinguishing microfossils from inorganic origin	[38]
Forensic science	Forensic image analysis review	[39]
Agriculture	Soil zymography	[40]
Forestry	Measuring canopy gaps	[41]
Viticulture	Robotic phenotyping of vine cultures	[42]
Manufacturing	Automated product inspection in mass production lines	[43]
Food science	Quality assessment of chicken meat	[44]
Engineering	Deformation measurement of cellular aluminium alloys	[45]
Material science	Damage evaluation after drilling laminates	[46]
Remote sensing	Analysis of anorthositic rocks in India as a model for lunar exploration	[47]
Water management	Monitoring of waste water true colour	[48]

Table 1.1: Some examples of the widespread use of computerised image analysis.

1.2.2 Advantages of Computerised Image Analysis

Image analysis, the area of biophysics this thesis forms a small part of, is also a part of many other subjects, as discussed above, and is an interesting field through its wide applicability and versatility.

Images have been a valuable form of data from the beginnings of scientific enquiry. From the development of lenses for magnifying small or far away objects to the development of specific labelling techniques and visualisation of otherwise invisible elements, the invention of new ways of seeing has often resulted in new insights and has led to new ways of looking at problems that had not been possible prior to that.

Up to only a few decades ago, the only way of analysing most images was through visual inspection by a human, whose experience and prior knowledge allowed him or her to draw information from the images and interpret them. The human brain through its

evolution has become very good at recognising patterns, but there has not been a need to develop similar skill in discerning absolute visual data, such as for example colour values or elements of frequency patterns, to name but two examples. Fig. 1.1 shows an illustration of the brain attempting to make logical sense of visual input at the expense of reporting accurate information.

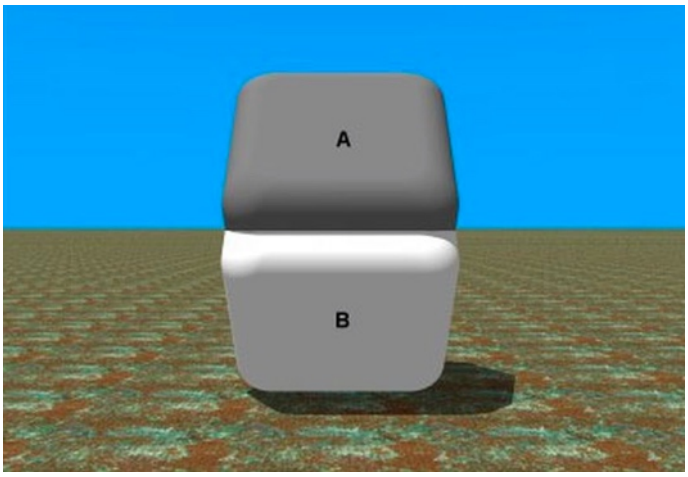


Figure 1.1: Illustration of the difficulty with obtaining reliable data through just looking. The shape above the horizon (A) seems of a uniform "grey" colour, whereas the shape below the horizon (B) seems "white". Now cover up the centre where the two areas A and B meet with your finger, and discover that A and B are in fact the same colour. (From [49])

Extracting data by visual inspection is prone to subjective bias and inconsistencies due to variations in the experimenter over time, when influenced by environmental conditions, health and emotional state, or by the activity the person has carried out just prior to the analysis. Images can contain a mix of relevant information and accidental features, which might interact to change the overall subjective impression, though not the relevant information content. The use of computers in extracting data from images can address many of those problems. Applying a consistent algorithm makes results from different images directly comparable, reliably reproducible, and eliminates the inconsistencies of human interpretation.

Computerised image analysis allows the extraction of features whose recognition is not readily available to the human brain, such as absolute values, value differences, frequency elements, the nature of a distribution or textural characteristics. In addition, often the accuracy of measurement by computer is greater than can be performed by a human, for instance when gradients in intensity or degrees of curvature are investigated.

Notwithstanding all this, the use of computers does not dispense with the need for human intervention. Researchers have to decide *how* to analyse an image, which features are meaningful and which are not, and develop algorithms that effectively measure and separate the two. In the end, a human being has to decide on the efficiency of the process

and the meaning of the outcome.

As discussed, images are intrinsic to many diverse areas of science, from the biological sciences to physics, environmental and surface science, medicine, geology and geography, astronomy and many more. Many of the images generated contain deeper levels of information than what is superficially accessible by looking. But overall the use of computerised image analysis is still patchy, although it has been around for many years. There may be several reasons for this slow adoption. In the past, the available hardware was in many cases limited to desktop computers with average specifications and cameras that took images on film. Knowledge of programming languages was not wide-spread amongst many scientists, and most computerised image analysis was rudimentary and required significant user input. Most image analysis was a measuring, rather than an analysis tool. Now a mounting awareness of the possibilities of computerised image analysis and the need to process larger data sets, together with the availability of more powerful hardware, digital photography and more accessible programming languages all pave the way for an increased use of these techniques. In some areas of science (such as physics) the use of computers and programming languages have been core skills for some time. In other areas, like for instance the life sciences, such skills are largely lacking. Also for this reason there is great scope for interdisciplinary collaboration between those scientists for whom such skills are daily routine and those for whom they are not, but whose images could benefit from data extraction with the help of computers.

Hence, computerised image analysis has the potential to be helpful to many researchers. It is applicable to any of scientific field, as long as image data are generated. It allows objective and reproduceable analysis of large numbers of images, and has the capability also to find kinds of information not accessible to the human eye. Obtaining more data from already existing images means a better return on the investment of time, money and effort needed to create them, and can enhance understanding of the system under investigation. If analysing images in this way were to become the standard, there would be great scope for advancement and benefit to many disciplines.

1.3 This Project

There are many kinds of PhD projects. Often they fall into one of two categories. The first is concerned with the investigation of a research question that, when answered, helps understand a wider problem. The question is addressed from different angles and using several appropriate experimental techniques. In this case, it is necessary to have a sound knowledge of several experimental techniques in order to interpret the results correctly, but a good overall knowledge of the system under investigation is paramount. In the second kind, the main focus is on an experimental technique, and various research questions, possibly in different systems, are addressed through this technique. In this case, a sound understanding of the experimental technique is paramount, while a complete background

of the research question is often confined to collaborators, and it may be sufficient to know in detail only the one specific experimental setup.

This project does not fall easily into either of these categories, since its focus is only on the development of ways of analysing images. The images we had access to do not constitute scientific data sets that can yield biologically important information, but are merely exemplary of such data sets. No experimental groups or treatment regimes were used, and the numbers of images were too low to achieve any statements of statistical significance. While we did not have access to such quantitative image sets, the techniques developed in this thesis are able to extract biologically relevant information from these kinds of images.

Image analysis is not a unified entity. As already mentioned, it is used in many fields, and depending on the image the approaches to analysis can be very different. From the application of relatively straightforward filters, and standard algorithms that are now inbuilt in many software packages, to highly specialised, complex and often not freely available algorithms related to problems like facial recognition or machine vision, the complexity of the approaches vary widely. This is reflected in the literature and in the fact that there are no dedicated journals to image analysis overarching the disciplines. This lack has the detrimental effect that developments in image analysis can hardly get published, unless accompanied by an experimental result. Details on image analysis are often buried in the methods sections of papers, without even a mention in the keywords. Searching for literature relevant to a problem in image analysis is therefore extremely challenging. Similarities to collagen might be stumbled upon in publications from food science, metallurgy, engineering, topology, nanofibre research, etc. While finding such publications often results from luck more than targeted searches, understanding these papers is another hurdle, since what was done often assumes a certain degree of prior knowledge in that field. On the theoretical side, mathematical image analysis papers often do not provide examples of applications and are not easily understood by non-mathematicians, due to the use of formalisms. Although there are a limited number of text books, many references to techniques are found in conference proceedings, which are hard to obtain and do not give full details. Thus, it is difficult to establish a core of knowledge of the current state of the art in image analysis, or even an overview of what has already been done.

It is hoped that works like this can make a small contribution to a field still very much in development, and that its potential may find wider recognition. This might result in the establishment of a dedicated journal, and a greater emphasis being put on these techniques by authors.

In this project algorithms have been implemented using the macro language of the freely available image analysis software package, Image SXM ([50]), developed by Dr Steve Barrett from the public domain image analysis software NIH Image ([51]). This

package was chosen because of the inbuilt functionalities and the easy availability of supervisory support. The aim was to extract information from complex and information-rich images of collagen, a biologically important and ubiquitous macromolecule. Images were taken with several different microscopic techniques common in the life sciences.

In the first chapter, we use images of collagen obtained by Atomic Force Microscopy (AFM), and develop an analysis strategy that measures both the thickness of one collagen fibre and the periodicity of its banding with very limited user interaction.

In the second chapter, Polarised Light Microscopy (PLM) images of collagen in thin sections are analysed by a textural analysis method we develop, and which yields three feature characteristics. Images from different groups can be distinguished through the combination of these, and this analysis is envisaged to have wider uses to categorise larger image sets.

In the third chapter the same analysis strategy is applied to images taken by Second Harmonic Generation microscopy (SHG) of the PLM samples. Results are compared with those of the previous chapter, and the robustness of the technique is confirmed.

Chapter four describes the image set obtained with Transmission Electron Microscopy (TEM) and the difficulties we encountered in their analysis. Furthermore, it describes a to our knowledge novel way of quantifying the success of analysis strategies and finding the optimal combination of analysis parameters.

Chapter 2

Collagen

2.1 Collagen – A very brief introduction

Image analysis begins with a specific kind of image, and with the consideration of what information should be extracted from these images. Considering the kind of images is important, as an image can only contain a selection of information filtered out from a more complex reality, but it is also important to be aware of as much information as possible about the object being imaged. As this work deals with collagen in various imaging techniques, it is helpful to review some basic facts about collagen and the sclera. Only a very brief description of collagen and its properties is presented here, as the field of collagen research is vast with over 200,000 articles published by 2015 ([52]) and only a small fraction of this information is likely to be necessary for image interpretation. For a fuller description of collagens see [53].

2.1.1 Why is collagen important?

Collagen is the most abundant protein in mammals, making up about 20 – 30% of the total protein mass ([53]). It is, together with water, the main constituent of tissues as diverse as tendon, ligaments, blood vessel walls, the cornea and sclera of the eye, skin, cartilage, bone, teeth and the fascia that surround all inner organs and muscle structures. Also other tissues, in which collagen is not the main protein, rely largely on collagen for their structural integrity, and also insects and plants use long fibrillar molecules for structural stability that, although not identical to collagen, are similar in overall structure and function.

Due to its pivotal role in the body, malfunction of collagen, either due to mutations of the genes, deficiencies in synthesis or processing, or alterations of the many accessory molecules that are required to produce a mature, functional collagen, lead to a variety of diseases, some of the most common are outlined here.

- **Scurvy** is one of the most well-known collagen-related diseases, which is caused by a deficiency in vitamin C. This vitamin is necessary for enzymatic processing of collagen's polypeptide chains in such a way that correct assembly can occur.
- **Ehlers-Danlos syndrome** is another example. It is an umbrella term for a relatively common set of conditions affecting around 1 in 15000 people. Mutations in different genes cause symptoms like hypermobility of the joints, weak muscle tone,

severe skin fragility, abnormal curvature of the spine, congenital hip dislocation, vascular problems and others. The severity can range from mild to life-threatening, as sufferers may be prone to sudden blood vessel or organ rupture. The genes affected are either that for collagen itself or proteins that process or interact with collagen, and thus almost all organs are affected in some way.

- **Osteogenesis imperfecta** is another group of diseases, which is caused by a different set of mutations, in which bones break easily, growth is stunted, there can be hearing loss, loose joints, breathing problems and other symptoms. The type of disease depends on the specific mutation responsible.

- **Scleroderma** is characterised by an increase in the production of collagen, which leads to hardening of the skin. The cause of this is unclear, and severity ranges from limited alteration of the skin in local areas to diffuse and systemic effects. In severe cases, people can die from heart, lung or kidney complications.

- **Alport syndrome** is caused by mutations in type IV collagen (for types of collagen see fig. 2.1). Here the function of the kidneys is severely compromised, and hearing loss and eye problems can occur.

- **Bethlam myopathy** is the consequence of mutations in type VI collagen, in which decreased amounts or formation of abnormal collagen type VI result in progressive muscle weakness and other symptoms.

- **Dermatosparaxis** is a genetic disorder predominantly but not exclusively found in cattle and sheep, where a deficiency in an enzyme required in the synthesis of collagen leads to extreme laxity and fragility of the skin and the internal membranes.

These are common collagen-related diseases, but the list is longer. Different collagens are not uniformly distributed amongst all tissues and can have very different functions, and therefore some mutations manifest themselves differently from others, but most of these diseases have at least a systemic component, reflecting the ubiquity of collagens. For a review of collagen-related diseases see [54] and [55].

Fig 2.1 shows examples of where in the human body collagen can be found in considerable amounts, however, it should be kept in mind that to some degree collagens pervade all tissues.

2.1.2 What makes a collagen?

There are many different types of collagen. The largest macroscopic collagen-rich tissues, such as tendons and ligaments, contain mainly type I collagen, which due to its abundance is relatively easy to extract and analyse. Its protein structure and sequence served as the model for the “classical” collagen properties. It consists of three helical polypeptide chains that combine to form a triple-helical, rope-like molecule. At both ends of this rope non-helical domains can be found that play a regulatory or functional role, for example regulating molecular assembly. Gradually, more structural collagens were identified that

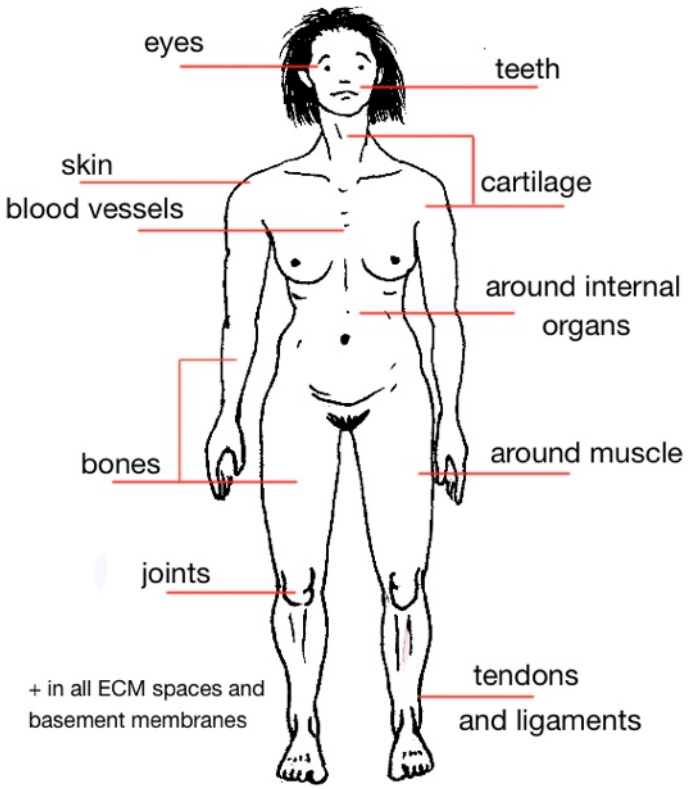


Figure 2.1: Illustration of the ubiquitous presence of collagen in human tissues. Major collagenous tissues are labelled, however, due to being an important constituent of all connective tissue, it pervades the whole body at small scales.

have similar gene and protein sequences to collagen type I and therefore contain a triple-helical domain, but are not identical and have different tissue distributions. This led to the definition of a collagen family of about 28 proteins (in humans) that share similarities in their structures and functions [53], [56]. More recently, more proteins with similar sequences have been found, and the collagen superfamily now contains over 50 proteins [53], [57]. They share the characteristic collagen sequence in at least some of their domains, but not all of these have structural functions, like the original collagens. It is likewise becoming clearer that the original collagens may also have functions other than mere mechanical and structural support. For a review of the different classes of collagen see [58].

Table 2.1 briefly illustrates the diversity of types and structures of collagens.

Classification	Collagen types	Supramolecular structure
Fibril-forming collagen	I, II, III	Striated fibrils
	V, XI	Striated fibrils, retain N-terminal regulatory domains
	XXIV, XXVII	Unknown
FACIT[a] collagens	IX, XII, XIV	Associated with fibrils, other interactions
FACIT-like collagens	XVI, XIX, XXI, XXII	Interfacial regions, basement membrane zones
Network-forming collagens		
Basement membrane	IV	Chicken wire network with lateral association
Beaded filament-forming	VI	Beaded filaments, networks
Anchoring fibrils	VII	Laterally associated anti-parallel dimers
Hexagonal networks	VIII, X	Hexagonal lattices
Transmembrane collagens	XIII, XVII, XXIII, XXV	Transmembrane and shed soluble ecto-domains
	Gliomedins, ectodysplasin	
Multiplexin collagens (Endostatin–XV and -XVIII)	XV, XVIII	Basement membranes, cleaved C-terminal domains influence angiogenesis
Other molecules with collagenous domains	XXVI, XXVIII	Collagenous domains in primarily non-collagenous molecules
	Acetylcholinesterase, adiponectin, C1q, collectins, surfactant protein, others	

[a]Fibril-associated collagen with interrupted triple helix

Table 2.1: Classification of collagen types by structure and function. Adapted from [58].

2.1.3 Molecular Collagen

Collagen is synthesised primarily in fibroblast cells[1] and excreted into the extracellular space. The steps necessary to make a functional, mature collagen molecule are numerous and complex. In common with all proteins, the genetic information is first transcribed from DNA into RNA (ribonucleic acid) format. This template is then translated into a polypeptide chain (so-called α-chains) by ribosomes (the machinery of protein synthesis in the cell).

The nascent chain enters the endoplasmic reticulum (a tubular network, in which further protein folding etc. occurs), where the first modifications of the chain and assembly into a collagen molecule (tropocollagen) starts. The molecule is then passed on

[1]Fibroblasts are the cells that make connective tissue.

to the Golgi apparatus (a system of vesicles that transport proteins to the outside of the cell) and is excreted from the cell via fusion of a Golgi vesicle with the cell membrane. During this process, post-translational modification continues and maturation is possibly not concluded until in the extracellular space. Occasionally, intermediary stages can be seen in TEM images, when cellular structures (fibripositors) appear to surround collagen molecules ([59]).

The repetitive motif of the fibrillar (fibre-forming) part of the molecule is an amino acid sequence of Gly-X-Y, where Gly is glycine, and X and Y stand for any amino acid, but are often proline and hydroxyproline, respectively. The hydroxylation of proline occurs as a post-translational modification through the enzyme prolyl hydroxylase and is essential for bonding between the α-chains. A lack of vitamin C, which is a cofactor for these enzymes, results in improperly formed collagens, which manifests as scurvy. One α-chain on its own forms a left-handed helix through its amino acid sequence. When three α-chains combine, they form a right-handed triple helix, in which the glycine residues of the α-chains are all oriented towards the centre of the rope, forming a hydrophobic core. This leaves the X and Y amino acids on the outside free to interact with neighbouring collagens and other molecules.

When α-chains are produced, they are longer than in the mature collagen molecule. This early form of unprocessed collagen is called procollagen (fig. 2.2). It is a common principle in protein synthesis that the primary gene product contains parts that are later cut out during post-translational modification of the polypeptide chain. Amino acids possess a carboxyl and an amino group, and proteins are synthesised in such a way that the amino group is free at the start of the chain (N-terminus) and the carboxyl group is used for polymerisation, and is therefore free at the end (C-terminus). In the case of collagen, various portions of the C- and N-terminal portions are cleaved off, but some sections before and after the helical segment remain (telopeptides). The extent to which this happens depends on the specific collagen. In the fibrillar collagens it appears that the C-terminal telopeptide is essential for initiating the twisting of the α-chains around each other, but the exact functions of the various N- and C-terminal sections are not yet fully understood. A collagen type I molecule at this stage is around 300 nm long and 1.5 nm in diameter. A highly ordered hydration coat surrounds the triple helix, and its role in stabilising the molecule is still debated.

2.1.4 Collagen fibrils

The fully processed and folded collagen molecules, in the case of fibrillar collagens, then assemble into collagen fibrils. Hydrophobic and electrostatic mechanisms are involved in the alignment, followed by stabilisation by covalent bonds via lysine and hydroxylysine residues. Collagen fibrils can vary in thickness depending on tissue and developmental age, and thicknesses from 50 to 500 nm have been found. It was found early in [61] and [62] that groups of five fibrils associate first (microfibrils) and then combine further to make larger fibrils, which contain both quasicrystalline [63] and more disordered areas. Reliably measuring the length of a mature fibril has been attempted, but not entirely achieved ([64]),

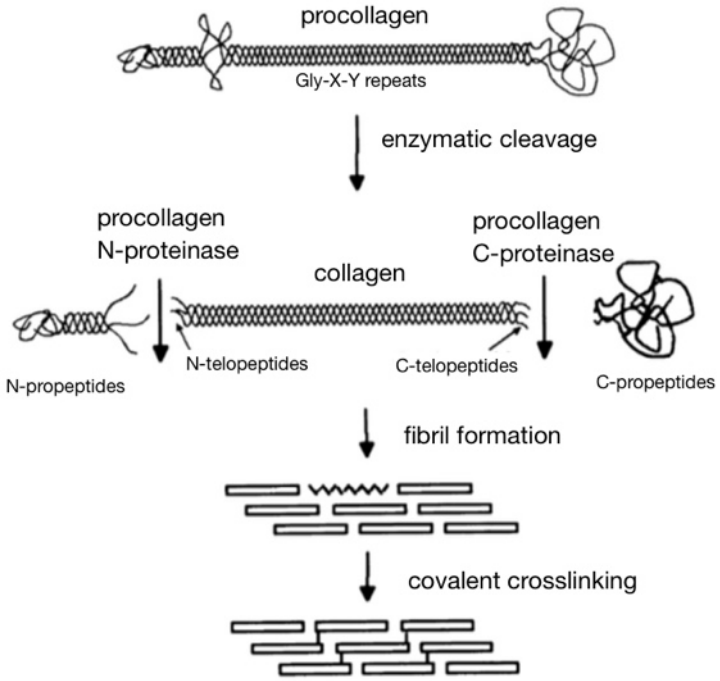


Figure 2.2: Processing of procollagen to collagen and assembly into fibril. Adapted from [60].

and it appears as if mature fibrils are no longer distinct, as during development, but are continuous within the large-scale fibre ([65], [66]). Fibrils can grow in circumference by addition of radial layers and longitudinally also by fusion of fibrils. It has been suggested that the assembly of collagen molecules into fibrils may be preceded by a liquid crystal-like arrangement that subsequently gets solidified into a fibril, but these studies have been in vitro and have now been superceded by studies showing the active involvement of fibroblast cells in steering collagen assembly and depositon [67]).

The regular arrangement of collagen into fibrils leads to a very specific and typical banding pattern observed in SEM, TEM and AFM. Between each collagen molecule and the next along the long axis there is a gap of about 40 nm. Neighbouring collagens parallel to the long axis are staggered with respect to their position along the axis by about 67 nm. This leads to a periodic pattern of density variation, which includes both an overlap and a gap region within every segment, and is called D-period. Fig. 2.3 shows how this periodicity arises. This is true for fibrillar collagens, and specifically for collagen type I. In the following, the results described refer to these types of collagen, as they are the most relevant in the context of this work, and also the ones on which most research has been done.

Although fibrils consist predominantly of one collagen type, other collagen types are almost always present. These minor collagens are thought to lie largely on the outside of the fibril, or contain portions that protrude outside, and could regulate fibril growth by preventing more collagens from being added. Type V collagen, for example, is often associated with type I collagen fibrils. This collagen has a large globular domain that protrudes

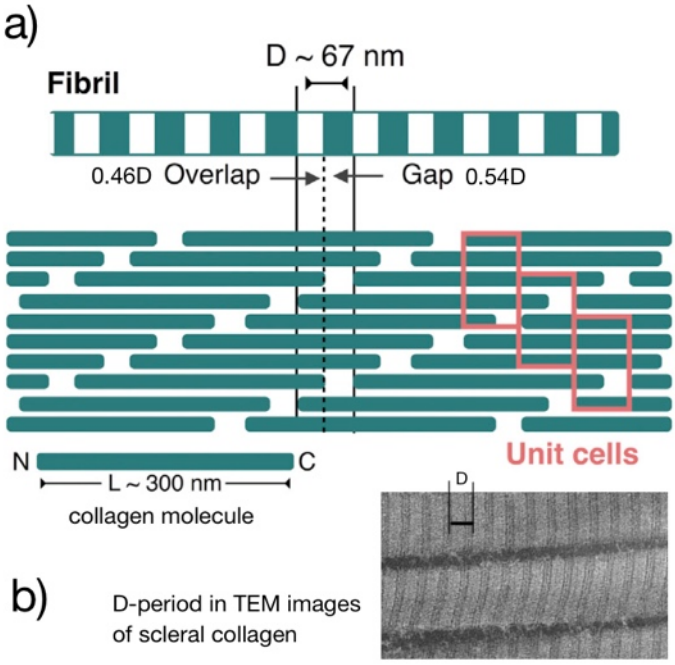


Figure 2.3: Staggered arrangement of fibrillar collagen leads to a periodic banding pattern in a fibril. a) Schematic drawing of the gaps and overlaps giving rise to a 67 nm D-period. b) TEM image of the same banding in a collagen fibril. Adapted from [68].

on the outside of the fibril. Within one fibril the C- and N-termini of the predominant parallel collagen molecules point in the same direction, but associated collagens can be in an antiparallel orientation. Fibrils can join with each other by lateral association of microfibrils and by fusion via their tapered ends, as long as their N- and C-termini point in the same direction ([53]). There is also some evidence that fibrils can be hollow on the inside and resemble tubes ([69]) in rat tail tendon, whereas other evidence point to a twisted rope structure ([70]).

2.1.5 Collagen fibres

In tissues, collagen fibrils associate to form collagen fibres, the organisation and diameters of which can be highly specific to the tissue and developmental age. For example, the different requirements of rhinoceros, cat, New Zealand White rabbit and chicken skin are provided by different arrangements of the collagen within the dermis [52]). In addition, Parry and Craig found also that distribution of fibril diameters are uniform and specific to the tissues they studied ([62]).

Different connective tissues require different responses to environmental challenges, and so the way fibrils are assembled into larger arrays varies in line with the biomechanical properties of these assemblies (fig. 2.4).

Proteoglycans such as decorin, biglycan, fibromodulin or lumican bind collagen fibrils together into fibres. The small core protein of those small leucine-rich proteoglycans (SLRPs) has specific binding sites for collagen, and the glycosaminoglycan side chains

(mainly chondroitin sulphate and dermatan sulphate) "decorate" the outside of the fibril. It is thought that inter-fibril interaction between these glycosaminoglycan chains, as well as interactions between FACITs (Fibril-Associated Collagens with Interrupted Triple helices) on the outside of the fibrils, are responsible for regulating the accretion and coherence of collagen fibrils and for playing an important part in mediating the whole tissue's response to tensile stress at the macroscopic level ([71]). A multitude of proteoglycans are found within fibrillar collagen structures, and interactions are complex ([72]).

The exact mechanisms of fibril assembly into higher order structures, as well as the role of fibroblast cells in these processes, are as yet not well understood. Morphologically, collagen fibres, surrounded by thin sheaths of connective tissue layers called endotenon, coalesce into small primary fibre bundles, these into larger secondary fibre bundles, and these into larger still tertiary bundles. Tertiary bundles make up a tendon. The tendons are ensheathed in a thin membrane called the epitenon. Within their fibres, the collagen fibrils are not only oriented longitudinally with the tendon axis, but also transversely and even horizontally, as well as in spirals [73]). This arrangement is thought to protect the tissue from damage due to mechanical forces from directions other than the main direction of tension. What happens in the fibrils and fibres when mechanical load is applied continues to be a subject for research. As biological tissues are anisotropic and individual, correlating overall behaviour with the contributions of subunits is difficult. Modelling is applied to investigate the mechanical response of the tissue. For example, Depalle et al. (2015) [74], adopting this approach, found a three-phase behaviour that they ascribe to i) uncoiling of the collagen molecule, ii) collagen molecules sliding against each other, and iii) stretching of the backbone of the collagen molecules, before rupture occurs.

2.1.6 Dynamic function

Historically, most research on biological samples involving images has been done on dead, and chemically altered (fixed) tissues. This has been unavoidable, because optical techniques for in vivo studies have not been widely available. This has been true on both the microscopic and the macroscopic scale. Adaptation of AFM to live cell imaging has improved the situation in some respects, but the mesoscopic scale has lagged behind. Recent observations by a plastic surgeon ([75], [76]), employing an endoscope during surgical procedures, has shed new light on the behaviour of living fascia (the connective tissue "envelope" of the body). His observations (fig. 2.5) suggest a dynamic flexibility and range of behaviours that have not been seen before, and are surprising, as these tissues had been considered inert for a long time.

This is an example of the importance of developing techniques that allow studying living subjects. New understanding on the systemic level can only up to a certain extent derive from knowledge of smaller and smaller isolated parts in a bottom-up fashion. Complementing such research with a top-down approach that records phenomena of a living organism in as complete a way as possible can give direction and new pointers to the study of the parts on the microscopic scale, even if, and possibly especially if, these phenomena

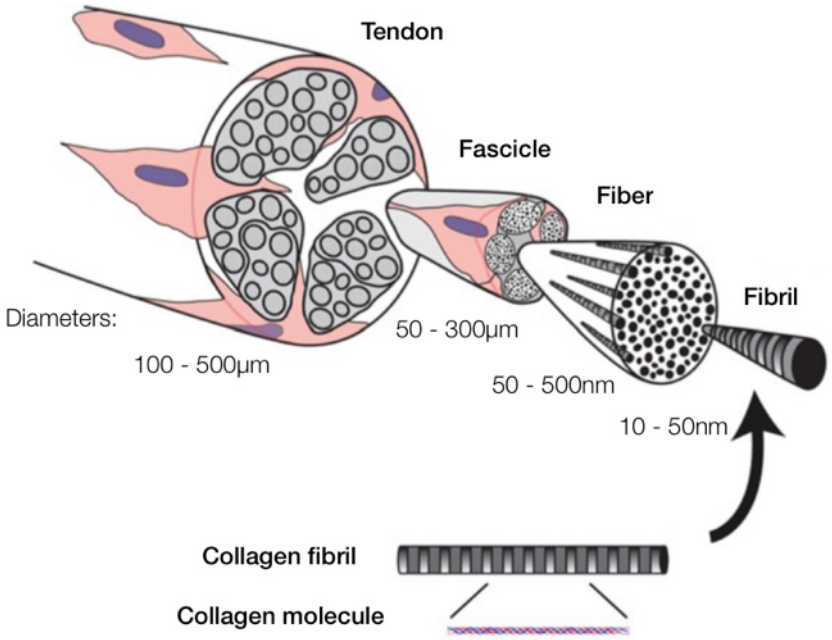


Figure 2.4: Schematic of the hierarchical assembly of collagen into tendon. Adapted from [58].

are not at all understood yet. New imaging techniques have often lead to leaps in understanding, and such novel recording techniques in living systems could be greatly helped by techniques in biological image analysis such as addressed in this thesis.

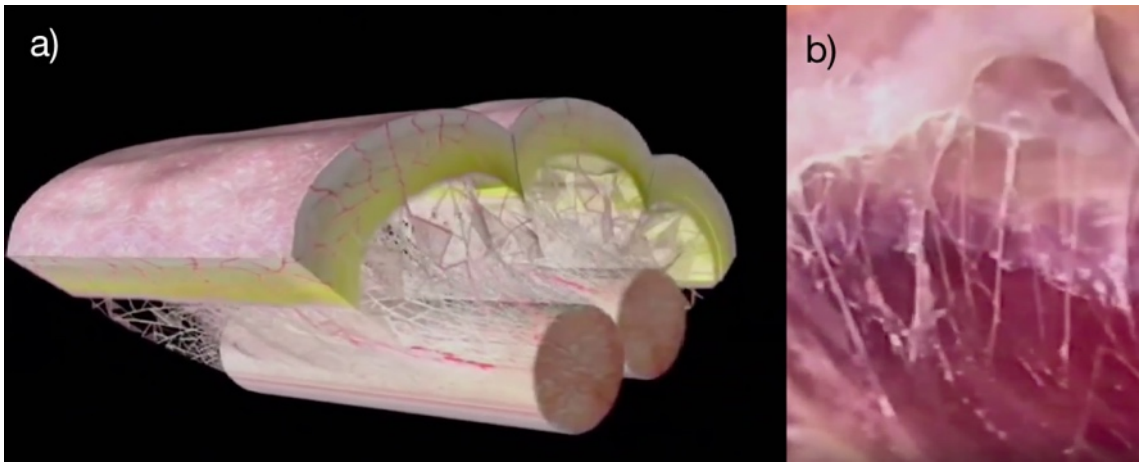


Figure 2.5: Multimicrovacuolar collagenic dynamic absorbing system (MCDAS), the ubiquitous, fractal-like structure that is proposed to envelop and permeate the entire body, from video "Strolling under the skin" ([77]). a) Schematic representation of the MCDAS around a tendon, b) endoscopy image of an MCDAS structure

2.2 Collagen in the sclera

2.2.1 Anatomy of the eye and sclera

The sample images that were available to us were all of collagen in the sclera of porcine eyes. Thus it is useful to briefly review the anatomy of the eye and discuss the collagen present there. As mammalian eyes are largely similar amongst different species, and most interest in ocular collagen research is focussed on treatment for human diseases, the anatomy of the human eye is described here (fig. 2.6).

The delicate contents of the eye ball are contained, and its whole structure is maintained, by the cornea and sclera. Both are connective tissue and consist predominantly of collagen, but the cornea is transparent and comparatively tender, whereas the sclera is opaque and tough. A much greater volume of research has been carried out on the cornea than the sclera, probably partly due to the mystery of its transparency, which is still not fully understood, and partly because up until recently the sclera was presumed to be a largely inert tissue. The sclera maintains the shape of the eye ball even when considerable forces are applied to it from the oblique and rectus muscles during eye movement. At its periphery, the cornea morphs seamlessly into the sclera, which continues to the back of the eye ball, where it merges with the dural sheath surrounding the optic nerve.

Inside the eye ball there are the aqueous humour of the anterior eye chamber, the iris and lens, and a number of smaller structures that aid in maintaining the tissues and generating clear vision. Behind the lens is the transparent globe of the vitreous body. At the back of the eye, as its innermost layer, is the retina with the retinal ganglion cells, which project their axons up along the optic nerve into the visual cortex. Below this layer, further to the outside of the eye, is the retinal pigment epithelium, which absorbs light and is essential for maintaining the retina. Below this is the choroid, which also contains pigments to absorb light, and blood vessels that provide nutrients for the retina below and the sclera above. Reflection of light on the choroid is responsible for the red eye effect in photographs.

Further to the outside of the eye ball begins the sclera. It is divided into four layers: the lamina fusca, the stroma, which is the thickest layer and can be regarded as the sclera proper, the episclera and Tenon's capsule (fig. 2.7, for review see [79]). The thickness of the sclera varies in different regions. At around 1 mm it is thickest near the optic nerve head and becomes thinner until it reaches a minimum of around 0.3 – 0.6 mm at the equator. Towards the anterior it thickens again and reaches up to 0.8 mm at the limbus.

- The **lamina fusca** contains the thinnest diameter collagen fibrils which intertwine and run in various directions. Many fibres extend into the choroid below and they can be wavy.

- Within the **stroma** there is a gradient of fibril diameters, from thin near the lamina fusca to thicker towards the episclera. In contrast to the cornea, where fibril diameter is both uniform at 25 – 33 nm ([64]) and in a very regular lamellar arrangement, in the scleral stroma fibril diameters can vary widely from 25 – 230 nm ([80]) and

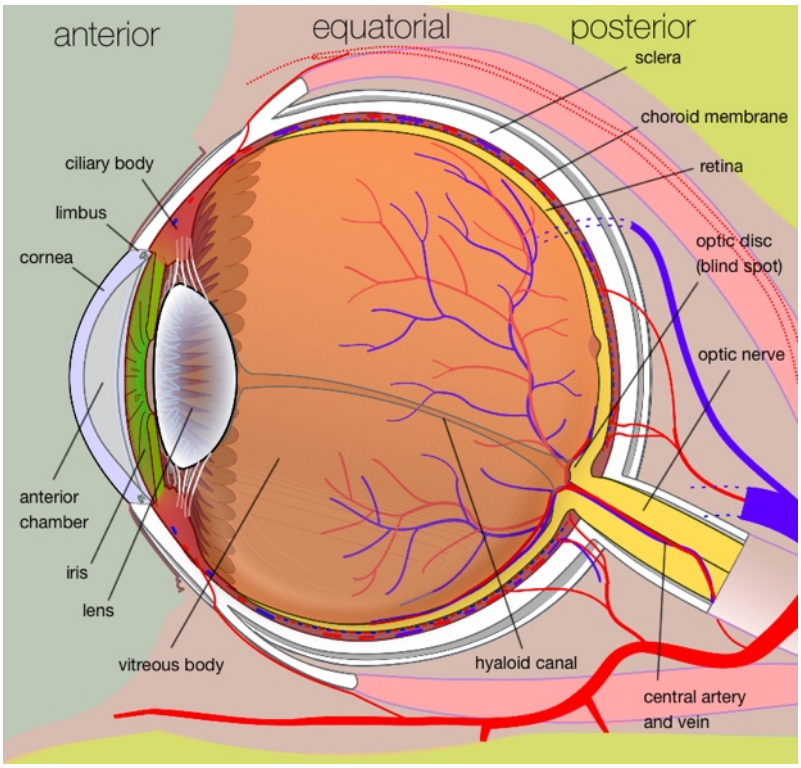


Figure 2.6: Anatomy of the human eye: only major landmarks are labelled. Adapted from [78].

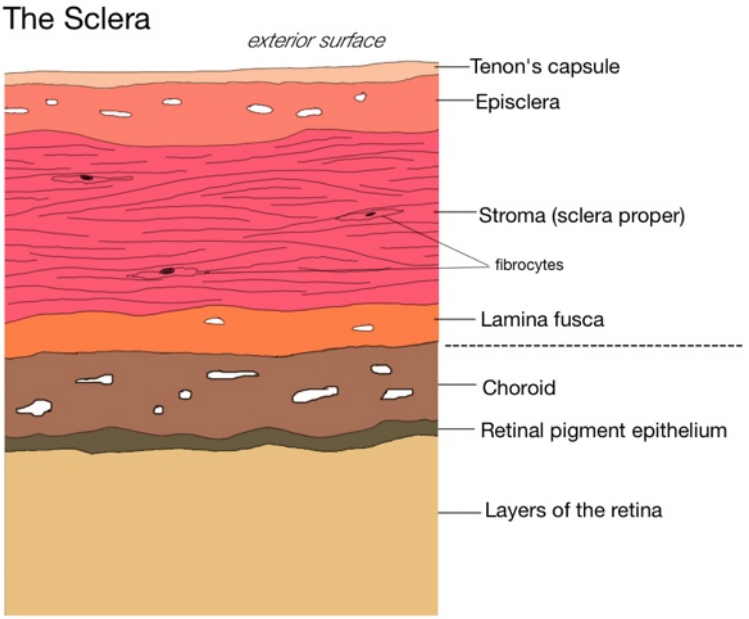


Figure 2.7: Schematic of cross-section through the posterior sclera. The dotted line separates sclera (above) from other structures (below) that are not part of the sclera but are drawn to provide context. Holes represent sections through blood vessels.

also the collagen bundles vary widely in size and branch and interlace extensively. Collagen bundles are smaller and more lamella-like on the outside, with widths of 1 – 5 μm and thicknesses of 0.5 – 2 μm, growing to widths of 10 – 50 μm and thicknesses of 1 – 6 μm towards the inside of the eye, although the collagen fibrils tend to be thicker towards the outside. Interestingly, the fibrils of the deep layers do not run in parallel with the bundles they belong to, but criss-cross in an irregular manner and interweave in three dimensions. A further constituent of the stroma are elastic fibres, fibrocytes and the extracellular matrix.

- The **episclera** is a thin and dense layer of collagenous tissue that unlike the stroma is well vascularised (supplied with blood vessels). It, together with the choroid, provides nutrients to the fibrocytes in the stroma, as there is no direct vascularisation there. The collagen fibres here are circumferentially arranged.

- In contrast to the lamina fusca, stroma and episclera, in which the transition is rather fluid, **Tenon's capsule** is a distinct, hypocellular layer of radially arranged, compact collagen bundles that only has trabecular (strut-like) links with the underlying episclera. It is firmly attached at the limbus and becomes more freely mobile towards the optic nerve, where it continues as a sheath around it. Where the rectus muscles attach to the eye ball it also ensheathes these and merges with their epimysium (the connective tissue envelope that surrounds muscle).

In the outer layers of the stroma, where collagen bundles show some directionality, it appears that the fibre direction is dependent on the local force distribution. It is a common motif for connective tissue, that can also be seen in bone and tendon for example, to develop local substructures that withstand the forces acting upon it.

The scleral stroma does not contain capillaries, but it is traversed by blood vessels that supply the choroid. Why capillaries do not grow in the sclera is unknown. Its nerve supply is however rich with nerve endings in the episclera and in the stroma around the blood vessel walls.

2.2.2 Collagen types and matrix

The main collagen in the sclera (and in the cornea) is type I collagen. It constitutes around 95 – 99% of the collagen present. The fibrils in the sclera are heterotypic, and other minor collagens are also present. In the sclera the main one of these is type III collagen, in addition to small amounts of types V, VI and XII ([81]). Marshall et al. [82] showed in study using immunogold labelling that types I, III and V are colocalised in fibrils, with type V being only at the fibril surface. Type VI collagen was only seen in the interfibrillar spaces.

Elastic fibres are especially abundant in the lamina fusca, but they are also scattered throughout the entire sclera. The interfibrillar matrix contains different glycoproteins and proteoglycans, the main kinds of which are decorin, biglycan and aggrecan. Decorin binds to the C-terminus of collagen type I in collagen fibrils and to type VI collagen in the interfibrillar space, and together with biglycan is thought to regulate collagen fibril

growth and interactions. Aggrecan plays a role in regulating tissue hydration, which may be one of the important differences between sclera and cornea, as the cornea is about 10% more hydrated and contains four times more proteoglycan, and the sclera becomes more transparent when it becomes dehydrated ([79]).

It is worth remembering that not all structure can be seen at all levels of magnification, and that some kinds of information require indirect analytic techniques. For example, fibre orientations may be seen by light microscopy, but the fibril dimensions are beyond the resolving power of light microscopy, but can be imaged with TEM using shorter wavelengths or with AFM. Determining collagen type from an image is generally impossible, unless a specific labelling technique is used. Immunogold labelling, for instance, uses antibodies specifically created to recognise a part of a particular collagen type. The antibody is then linked to a gold particle and incubated with the tissue. When this is followed by TEM, the electron-dense gold particles can be found at the sites where the antibody has found its binding site. Thus, dark globules mark the site of a specific molecule, even if the molecule itself is invisible. Some molecules may be able to be identified in transmission electron micrographs, but usually only when they have very distinctive features and have been specifically identified previously. The same is true for AFM, but in this technique it is even more challenging to add specific labels.

2.2.3 Scleral disease

The human sclera reaches its maximum flexibility at the age of 12 or 13 years. Thereafter its rigidity increases and the gradient in fibre diameters through the sclera becomes more pronounced. Near the muscle insertion points the sclera becomes thinner and the tissue yellows as fat globules are deposited between the fibrils. This makes the sclera less resilient to mechanical challenges, and injury may be both more likely to occur and more difficult to repair.

Scleritis is an inflammatory reaction within the sclera. It is characterised by areas of tissue necrosis surrounded by inflammatory cells and is most often accompanied also by an inflammatory reaction in the surrounding blood vessels. Many factors can trigger this reaction, from surgery to systemic connective tissue diseases to infections, and the relative isolation of the sclera from the turn-over of body fluids facilitates long-lasting immune reactions. The process starts by activation of fibrocytes, which then appear to break down the proteoglycans which bind collagen fibrils together. The collagen fibrils then unwind and break apart, and become digested by cells in the vicinity. Collagen fibrils in areas where this process has been put in train appear swollen and unravelled. Although no common antigen has been identified, it is possible that the exposure of “naked” collagen outside the confines of a fibril to the immune system could act as an antigenic stimulus and exacerbate the situation by instigating a positive feedback loop ([79]).

Myopia, or short-sightedness, is a very common condition in which the focal point of the optic system is located in front of the retina. This is mostly due to an elongation of the posterior eye ball. If this is extreme, the retina can degenerate and become detached as a consequence. It is not yet clear what causes myopia, but it is likely not one single

factor, as in some cases there seems to be a genetic component, and myopia can be induced in an animal model by light occlusion. The sclera shows a two-phase response to deformation. At first, there is a short phase of quick, elastic lengthening, followed by a long phase of slow stretching. The deformation comes mainly from the posterior eye, which is more extensible than the anterior or equatorial portions. The stress-strain curve shows a considerable degree of hysteresis, so that when the force is removed, the eye ball returns to a similar but not quite identical conformation to the starting shape. Also under constant force in the same range as physiological intraocular pressure the eye ball first shows elastic expansion followed by a slow creep, the rate of which depends on the rate of axial elongation. It is interesting to note that in the early stages there can be a tissue loss up to 7% ([83]), which might indicate an active process. Although the sclera undergoes constant remodelling throughout life, the turnover time is long and in myopia the posterior sclera is significantly thinned, which might suggest that the mechanism by which the eye elongates is stretching rather than active tissue remodelling. In severe myopia the gradient of fibril diameters through the sclera is lost and only thin fibrils can be found, which would, in turn, suggest some tissue remodelling. Although whether this is a primary mechanism or a later adaptation is debatable. It is likely that the loss of tissue mass and the growth in volume of the eye ball are caused primarily by a change in proteoglycans, perhaps via changing the ease with which collagen fibrils can slide past one another, as changes in the appearance of collagen can only be seen some time later.

What ageing, scleritis and myopia all have in common is a more or less dramatic change in collagen fibril and fibre morphology and arrangement. Thus, to study any of these conditions further, quantitative methods of estimating collagen characteristics on various length scales are likely to be invaluable tools. Furthermore, mechanical modelling of scleral behaviour is used for instance in myopia research, but also in other disease states, to try to understand the contribution of collagen fibres to the development of, or resistance to, disease. Modelling approaches will also benefit from an input of data on collagen distribution and organisation derived from automated image analysis methods of the relevant tissue areas.

Chapter 3

Atomic Force Microscopy

3.1 Atomic Force Microscopy

3.1.1 Principles

Atomic Force Microscopy (AFM) is one of a number of Scanning Probe Microscopies (SPM) that were pioneered by Binnig and Rohrer in the 1980s ([84]). Such was the impact of this technology that the inventors were awarded the Nobel prize for physics for this work in 1986. SPM is a mode of microscopy that allows magnification far below the scales resolvable in light microscopies, and is based on physical detection of surface features such as profile, electrical conductance, charge, stiffness, etc. These properties are measured by the interaction of a sharp probe with the sample surface (see below).

While SPM techniques do not allow to "see through" biological tissues, like light-based microscopies, they have nevertheless revolutionised biological research because of the ability, especially of AFM, to observe biological specimens without any preprocessing (such as chemical fixation methods, thereby reducing the likelihood of artifacts) in air and even under liquid, at room temperature. Also, greater resolution is possible than with conventional light microscopy.

In vivo measurements on biological specimens in more natural aqueous environments can yield biologically more relevant results than on dehydrated and fixed specimens, which are common in light microscopy. Many biological samples retain their native structure and function much better under these more life-like conditions. The ability to examine biological specimens at high resolution, without preprocessing, at room temperature and under liquid is probably the greatest advantage of AFM over conventional light microscopy.

3.1.2 The Atomic Force Microscope

The Atomic Force Microscope (AFM) consists, essentially, of a sharp tip mounted on the underside of a flexible cantilever held a short distance away from a piezo-controlled stage on which the sample is mounted (fig. 3.1). In the simplest imaging mode, the tip is brought closer to the sample surface until an interaction becomes measurable. A laser is directed onto the upper, reflective side of the cantilever and the laser beam is reflected onto a photodiode, which registers deflection of the cantilever due to the interaction of the tip with the surface. As the tip scans across the sample surface, a computer-controlled feedback mechanism re-adjusts the piezo elements of the sample stage in such a way

that a constant force is maintained. The sequence of these adjustments are then used to reconstruct an image of the sample surface line by line.

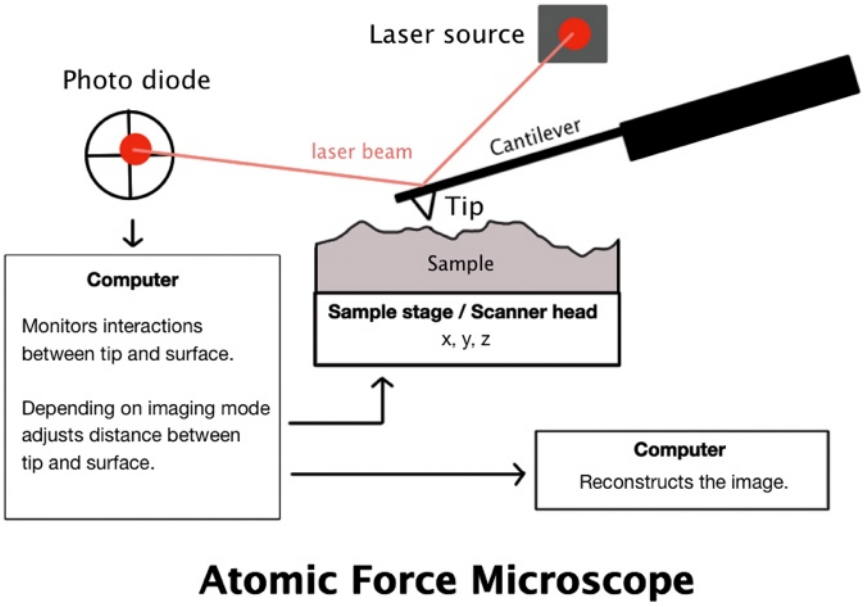


Figure 3.1: The principle of the Atomic Force Microscope. A laser beam detects the deflection of the cantilever as the tip scans over the sample surface. A feedback loop readjusts the tip-sample distance and an image of the surface is reconstructed.

3.1.3 Imaging Modes and data interpretation

Contact mode

There are numerous ways in which the principle of AFM can be employed to investigate a surface ([85]). Contact mode is historically the earliest, and the simplest form of AFM imaging. The AFM tip is dragged across the sample surface and the interaction forces between it and the profile of the surface bend the cantilever up and down, which results in a deflection of the laser beam shining onto the back of the cantilever. This deflection is picked up and measured by a photodiode divided into quarters. A feedback loop maintains the cantilever deflection at a constant load force. The surface profile is then reconstructed from the movement of the z-piezo element that is necessary to maintain the load force. The resulting image shows the height profile of the sample surface.

This imaging mode has several drawbacks. Firstly, the lateral forces involved in dragging the tip along could disturb or destroy soft surface features and potentially damage the tip. Secondly, when the sample surface is disturbed, parts of it may become dislodged and be dragged along with the tip, which alters tip geometry and leads to faulty image reconstruction. Thirdly, the load force corresponds to a specific cantilever deflection only if the surface properties stay constant. Thus, it is possible that difference in material characteristics across a scanned surface could mimic a height difference, which complicates the

interpretation of the images. Furthermore, features that are smaller or sharper than the tip used are reconstructed as a mirror image of the tip (convolution of volumes). This is caused by the bulk of the tip sliding along such a surface feature with each point on the tip describing the shape of the tip (fig. 3.2). Other tip artifacts are caused by contaminated tips where tip geometry is altered by debris picked up from the surface, or blunts tips where contact with the surface has eroded part of the tip. Contaminated tips may also cause an artifact called "multiple tips" where particles clinging to the tip act as additional contact points and produce double or multiple images of the surface features. Further artifacts can be caused by inadequate hardware performance of the AFM microscope.

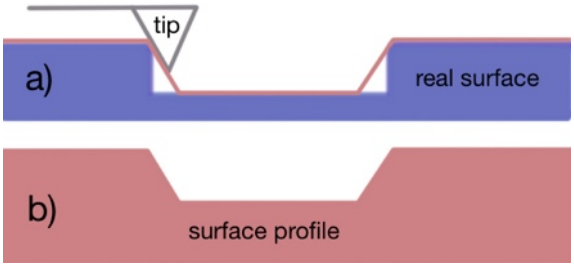


Figure 3.2: Convolution of volumes, a common tip artifact. The AFM tip is less sharp than a surface feature and it is the tip geometry rather than the surface geometry that constrains the path. a) Tip scans across real surface (blue) and describes a path (red line); b) The reconstructed image shows mirroring of the tip geometry. The slopes are artifacts of the tip geometry.

Intermittent contact mode

In order to address the problem of sample and tip damage due to lateral shear forces, the next stage of development in AFM imaging was an intermittent contact mode, also called dynamic mode, in which the forces between tip and sample are much weaker and therefore allow imaging more fragile structures ([85], [86], [87]). The first implementation of this mode was amplitude modulation, in which the tip is driven by an oscillator at a constant frequency slightly above or below the resonance frequency of the cantilever ([88]). (Attractive forces decrease the resonance frequency of the system, while repulsive forces increase it. Choice of driving frequency is guided by the nature of the surface and the conditions in the experiment.) The force gradient between the tip and sample surface is changed by the surface profile, and the mechanical or electrical properties of the surface. This leads to a change in the resonance frequency of the system, causing a change also in the amplitude of the tip oscillation and a phase shift relative to the driving signal. These are detected as the feedback signal.

Since the reaction time in this mode is relatively slow, an improvement was developed by Albrecht and colleagues ([89]) in the from of frequency-modulation. In this mode the driving frequency is kept exactly at the resonance frequency. As the tip-sample interaction changes the force gradient between them, a feedback loop keeps the driving frequency at

the resonance frequency, and the amplitude remains constant. This mode of operation is faster than amplitude modulation.

Both these dynamic modes operate at a distance relatively far from the surface, where attractive forces between tip and sample predominate. This distance, however, also includes the layers of adsorbed water that covers the surface at ambient conditions. Contact mode relies mainly on repulsive forces and operates at a distance where the tip is trapped between the surface and the meniscus of the adsorbed water layer, which can lead to sticking of the tip. In response to these limitations Zhong and colleagues developed "tapping mode" ([87]), which is a mode operated at close distances in the region of repulsive forces and allows successful imaging under ambient conditions. This mode is very widely used today in general, and particularly in biology. It may be used as an amplitude-modulated or a frequency-modulated method. It forms the basis for the development of Bruker's PeakForce QNM (Peak Force Quantitative Nanoscale Mechanical Characterisation).

In Bruker's PeakForce QNM ([90]) a force curve is taken at each point of the image scan. A force curve is a plot of the forces exerted on the tip as a function of distance of the tip from the sample surface. In practice, the y axis shows the peak force error, the difference between the actual peak force on the tip and the set point peak force (fig. 3.3). The forces involved are both attractive and repellent, and consist largely of van-der-Waals, electromagnetic and capillary forces ([90]). At a long distance from the surface, only long-range forces affect the tip, which are very small. As the tip approaches the surface, attractive forces increase and eventually these forces overcome the cantilever stiffness and the cantilever bends downwards. As the tip gets pulled closer to the surface, the repulsive forces increase until the tip reaches its minimal distance to the surface. At this point the repellent force is strongest, and is thus called the "peak force". In Peakforce QNM it is usually the amplitude of the tip oscillation that is kept constant by the feedback loop and is used to reconstruct the topographic image. When the tip is withdrawn, the attractive forces increase again until a point is reached where the tip is eventually pulled off the surface again and returns to a distance where small, long-range forces are the dominant influence. Whereas force curve measurements are taken also in contact mode to estimate the overall tip-sample interactions, this is normally only done on selected points, as it has to be set up independently of the scan. Peakforce QNM is now able, however, not only to obtain a force curve at every scan point, but also to analyse the curve in real time. This not only allows the use of the peak force for controlling the feedback loop in a way that is responsive to the local sample properties, but it also allows the reconstruction of images of a range of mechanical properties derived from properties of the force curve. The standard images derived from force curve properties are the Derjaguin-Müller-Toporov (DMT) modulus, adhesion, dissipation and deformation. See fig. 3.3 for illustration, and ([90]) for a review of the Peakforce QNM principles described above. Additional modes that use the raw data of the force curve for calculations using different models can also be added, replacing some of the defaults. In all AFM imaging modes, assumptions about tip geometry play a crucial role for how the images are reconstructed. Different assumptions on this lead to different theoretical models. For example, the DMT model is based on the

assumption of a spherical tip, whereas the Sneddon model ([91]) assumes a conical tip, which is sometimes assumed to be closer to the real interaction in soft samples, where the tip indents.

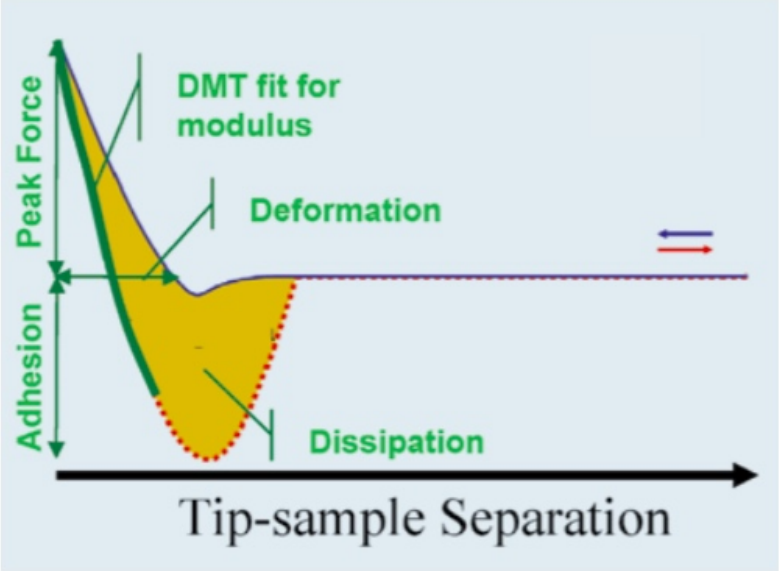


Figure 3.3: Typical force-distance curve in Peakforce QNM. On the curve the parts used for the different default channels of Peak force error, DMT modulus, adhesion, dissipation and deformation are labelled accordingly. Modified from [90]

3.2 AFM Images and Analysis

3.2.1 AFM Images

The images used for analysis were kindly provided to us by Dr Riaz Akhtar (Department of Mechanical Engineering, University of Liverpool) and are of porcine sclera. This tissue is full of collagen bundles of varying sizes and thus provides good samples for the study of collagen. The eye samples were snap-frozen and then cryosectioned before being imaged by AFM in air.

These images contain the following 8 channels (see fig. 3.5):

- Height
- Peak Force Error (PFE)
- DMT Modulus
- Log DMT Modulus
- Adhesion
- Deformation
- Dissipation

- Sneddon modulus

A limited number of images were used for analysis, because they were of sufficient quality and contained collagen fibres in a clearly recognisable and more or less linear arrangement. See fig. 3.4. Only the height and peak force error images are shown. For examples of the remaining channels, refer to figs. 3.5 and 3.8. Image labels were set by the experimenter and hold no significant information in this context. The images used are shown in fig. 3.4. All images are squares of 256 x 256 pixels.

3.2.2 Biological Significance

In the eye, and specifically in the sclera, collagen, as a predominant tissue component, is largely responsible for the mechanical and structural integrity of the eye ball. As mentioned previously, how and in what way deformation of the eye occurs in disease states is closely related to the way collagen is deposited and synthesised there ([92]).

Also in a wider biomedical context, biological images such as these may be used either for research purposes or in a clinical setting. Many diseases are linked with abnormal formation or degradation of collagen, leading to atypical appearance in tissue samples. To name just a few examples, in osteogenesis imperfecta collagen type I is insufficient or improperly formed, and Ehlers–Danlos syndrome comprises a group of genetic disorders characterised by defects in the structure, production, or processing of collagen. Furthermore, there are dietary conditions which affect collagen, such as lathyrism, a copper deficiency, in which the enzyme lysyl oxidase is inhibited, leading to a lack of collagen crosslinking, or scurvy, a deficiency in ascorbic acid (vitamin C) required for the formation of hydroxyproline, resulting in blocked collagen formation. It is also thought that increased cross-linking of collagen in ageing contributes to increased brittleness of skin and bones, but the mechanism for this remains hypothetical ([93], [94]).

It is likely that analysis of collagen fibre thickness and banding periodicity in AFM images could be used in studies aimed at further understanding these conditions.

3.2.3 Analysis strategy

Two common characteristics of collagen that have biological relevance are the periodicity of the banding and the fibre diameter. Periodicity of banding is a direct result of the self-organisation of collagen triple helices into fibres, and can vary in different tissues, according to hydration state or in disease states ([53], [95], [96]). It is therefore a useful quantity to investigate. Fibre diameter also varies widely and is, among other factors, responsible for the mechanical strength of a tissue. It also varies with hydration state and age of the tissue ([97], [98]). Fibre diameter and periodicity are both universally interesting and most easily accessible from a visual inspection of collagen images. Hence, the purpose of the analysis strategy was aimed at extracting information on collagen fibre diameters and banding periodicity from the images.

2E.010
Dimensions 1.48μm $\times$ 1.48μm
Height z-scale 0-140 nm
PFE z-scale 25.9mV

2A.010
Dimensions 1.51μm $\times$ 1.51μm
Height z-scale 0-123 nm
PFE z-scale 20.2mV

2PA.034
Dimensions 1.40μm $\times$ 1.40μm
Height z-scale 0-117 nm
PFE z-scale 30.0mV

3A.009
Dimensions 1.39μm $\times$ 1.39μm
Height z-scale 0-189 nm
PFE z-scale 25.5mV

3EB.035
Dimensions 1.39μm $\times$ 1.39μm
Height z-scale 0-96 nm
PFE z-scale 42.4mV

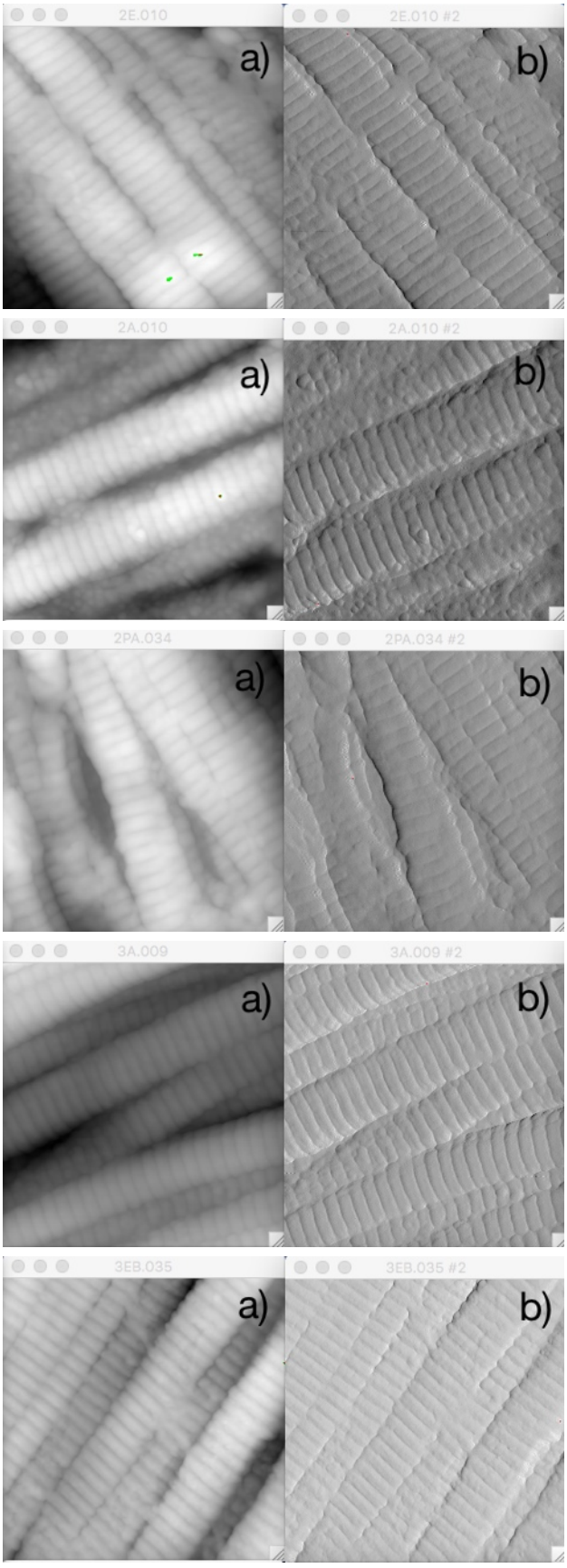


Figure 3.4: Overview of height (a), and peak force error (PFE) (b) channels of the images 2E.010, 2A.010, 2PA.034, 3A.009 and 3EB.035 used for algorithm development. All images are 256×256 pixels.

Due to the already mentioned advantages of an automated or semi-automated approach to image analysis, it would be advantageous to develop a strategy that would extract that information with minimal risk of bias or of error due to human variability.

The eight channels of the AFM images show marked visual differences. As image analysis works with the data that make up the image, i.e. the grey (or RGB) values of a set of pixels, it is less important how the data in the image is obtained and what the image data are, as long as there is a direct correlation between the grey values (gvs) in the image and the underlying feature that has given rise to them. For example, if the image of the adhesion channel shows good discrimination between fibre edges and fibre, then this image could be used instead of the height image, which is expected to depict the "real" fibre edge, which is the topographic dip between one fibre and the next. (Compare a) and e) in fig. 3.5.) It is obvious from first visual inspection that not all channels are equally useful in depicting our characteristics of interest (fig. 3.5).

Human perception is not always good at seeing the small details, but can often quickly grasp the overall pattern. On occasion the visual system even "invents" patterns where there are none, or can find subtle patterns within a noisy environment. Thus, a human observer can sometimes "clearly recognise" distinct fibres, when on closer inspection individual fibres are not distinguishable. This ability to recognise patterns is very hard for a computer to emulate. An easy indication of how useful these channels will be for finding fibre edges and banding is to plot a profile across the relevant features.

Figures 3.5, 3.6 and 3.7 show a comparison between the line "Regions Of Interest" (ROIs)[1] drawn across fibre boundaries in the different channels, and the corresponding plot profiles. Figures 3.8, 3.9 and 3.10 show the equivalent for line ROIs drawn along the fibre axis to show collagen banding.

From figs. 3.6 and 3.7 it is obvious that a and b (height and peak force error) yield the clearest plot profiles for detecting fibre edges.

In the case of the banding, fig. 3.9 and 3.10 show clear periodicity in a, b, c, and d, a less clear but still recognisable pattern in e and g, and considerable noise in f. In h all periodicity is lost.

Thus, height (a) and peak force error (b) yield the clearest plot profiles for the detection of both fibre edges and banding, which means that they can be used in both analyses. The remaining channels will therefore be disregarded for the purpose of this analysis.

The quantities that should be analysed, periodicity of banding and fibre diameter, can be manually measured by drawing a line ROI across the image, making a profile plot of the ROI and reading off the distances. This becomes less feasible if a large number of images are to be analysed and if experimenter variability is to be avoided. Ideally, an automated approach should handle everything by itself, from opening images in a folder, to selecting the right image characteristics, analysing and measuring them, calculating or

[1] "Region Of Interest", or ROI, is a term used in image processing and analysis for any selection made in an image. This can be a line or an area, such as a circle, rectangle or random shape. When this selection is made, any subsequent operation is done on this part of the image only.

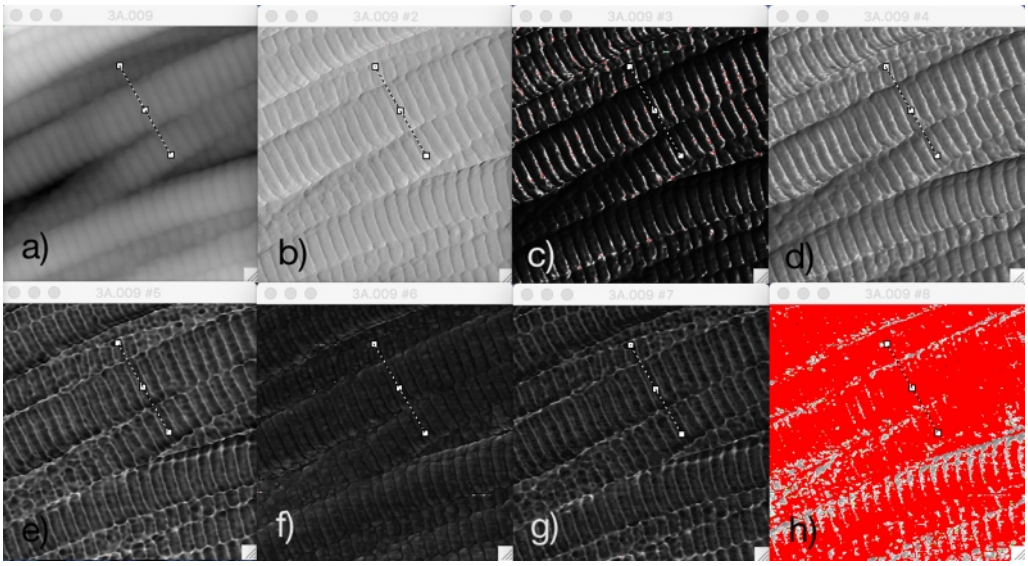


Figure 3.5: An example AFM image with its 8 channels. An identical line ROI has been drawn in each channel to show fibre edges. a) height; b) peak force error; c) DMT modulus; d) Log DMT modulus; e) adhesion; f) deformation; g) dissipation; h) Sneddon modulus. The corresponding plot profiles are shown in figs. 3.6 and 3.7.

collating the numerical data, and saving them.

While initial attempts were made to generate such an all-encompassing, autonomous strategy, it soon became obvious that this wasn't feasible with the limited scope the macro language in Image SXM provides ([51]). The main challenge proved to be the complexity of the image structures. Collagen fibres do not always run parallel or stay in the image plane, but appear and disappear underneath and above each other, and often, while the eye might be able to deduct fibre edges from the overall impression, actual edge features appear and disappear along the fibre edges, too. An example of this is shown in fig. 3.11 where fibre edges and features that might be identified as such were manually traced, as they could be identified by eye. What this reveals is that while the eye is able to identify fibres as a whole quite well, the actual boundaries are not always necessary for this and are often absent.

In addition, scanning artifacts can mimic fibre edges as well as banding lines, and tissue structures other than collagen may be present and obscure parts of the image (not shown).

For an algorithm to be able to distinguish fibres that are suitable for analysis, it is likely that some sort of "training", for instance of a neural network, might be required, which goes beyond the scope of this project, and would likely also go beyond the scope of most image analysis users, for whom a more straightforward approach would be better suited.

One possible solution to the problems arising from the diversity of fibres might be to employ a whole-image or textural approach, where the overall direction of fibres, overall periodicity and overall distance between fibre edge-like features might be analysed. This would have drawbacks, however. The nature of the data obtained this way would be slightly different from single fibre measurements. No reliable information would be gained

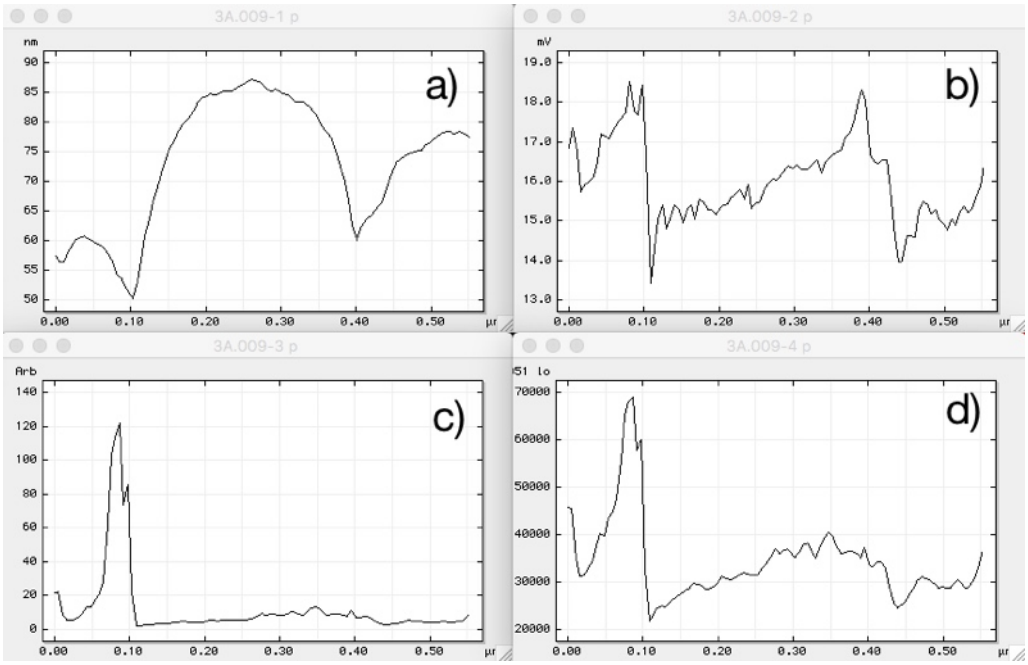


Figure 3.6: Plot profiles of channels 1 – 4 (a – d) in Fig.3.5. a) height; b) peak force error; c) DMT modulus; d) Log DMT modulus.

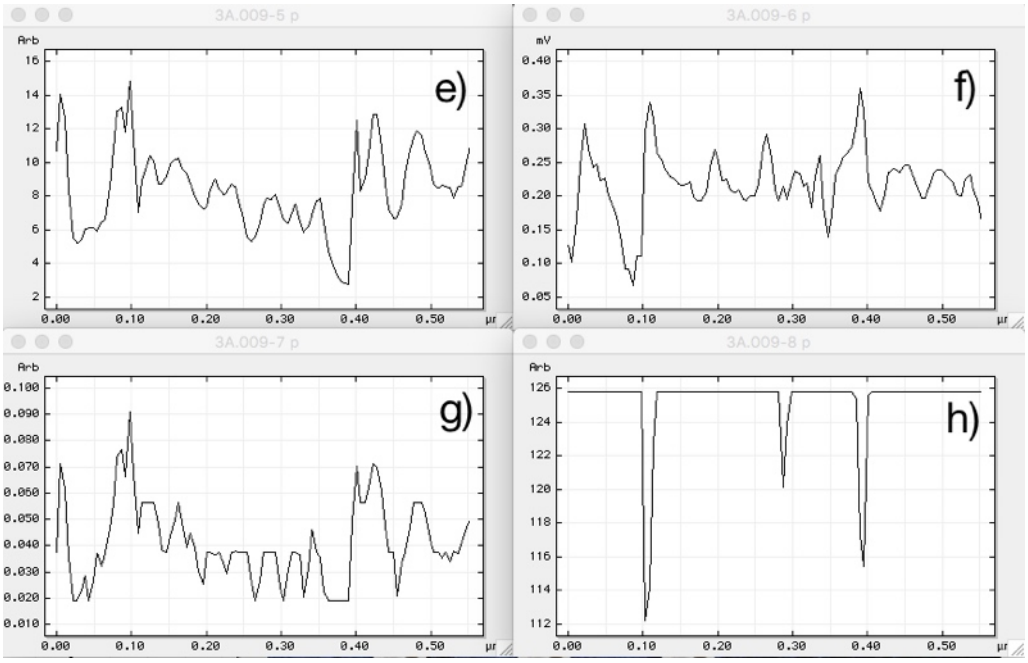


Figure 3.7: Plot profiles of channels 5 – 8 (e – h) in Fig.3.5. e) adhesion; f) deformation; g) dissipation; h) Sneddon modulus.

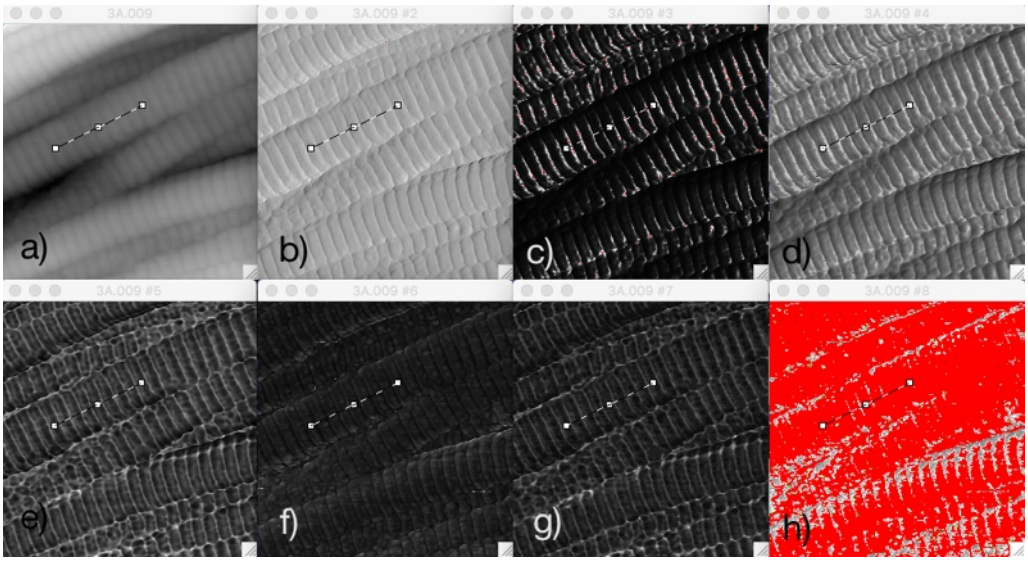


Figure 3.8: Another example AFM image with its 8 channels. This time, an identical line ROI has been drawn in each channel to show banding along the fibre axis. a) height; b) peak force error; c) DMT modulus; d) Log DMT modulus; e) adhesion; f) deformation; g) dissipation; h) Sneddon modulus.

about individual fibres, and the overall value could be affected by many hard to control factors. The problem of separating measurable from non-measurable areas would remain, as the inherent error of the measurements would depend on what proportion of the image contributes to the measurement values, and which other features that should not have been included did in fact contribute to the measurement. It was thus concluded that a textural approach was not appropriate in this case.

It was decided that it was acceptable to have an initial user interaction for each image, in order to eliminate the problem with selecting image areas suitable for analysis. It is comparably easy for a human to see which fibres lend themselves to analysis and to draw a line ROI near the centre of such fibres. Although this produces a bias towards fibres that are easily measured, which may not be a purely random selection of the whole fibre population, this problem would also be intrinsic to manual measurements. Any analysis of differences in fibre diameters and periodicity between experimental groups would have to keep in mind the fact that some fibres are harder to measure than others and take appropriate steps to investigate whether this influences the results.

3.3 Macro Analysis

3.3.1 Macro outline

Following the considerations outlined above, the theoretical analysis strategy was translated into an algorithm in macro language. A simplified outline of the algorithm is shown in fig. 3.12.

Briefly, the user draws a line ROI into an image. The macro is then started and first makes a copy of the active image and the ROI. (This is done so that alterations

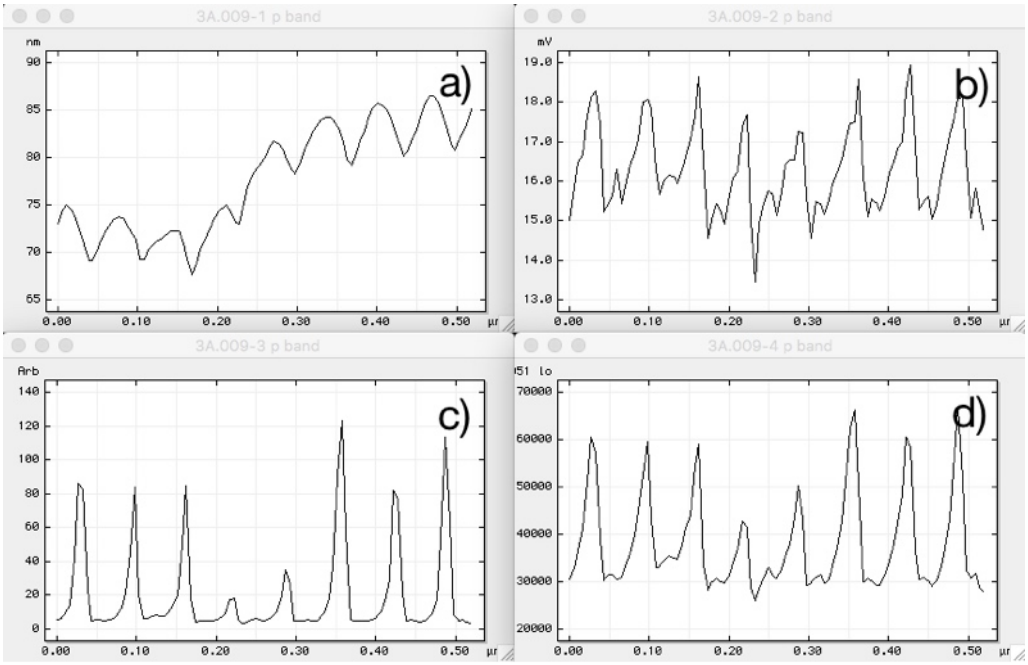


Figure 3.9: Plot profiles of channels 1 – 4 (a – d) in Fig.3.8. a) height; b) peak force error; c) DMT modulus; d) Log DMT modulus.

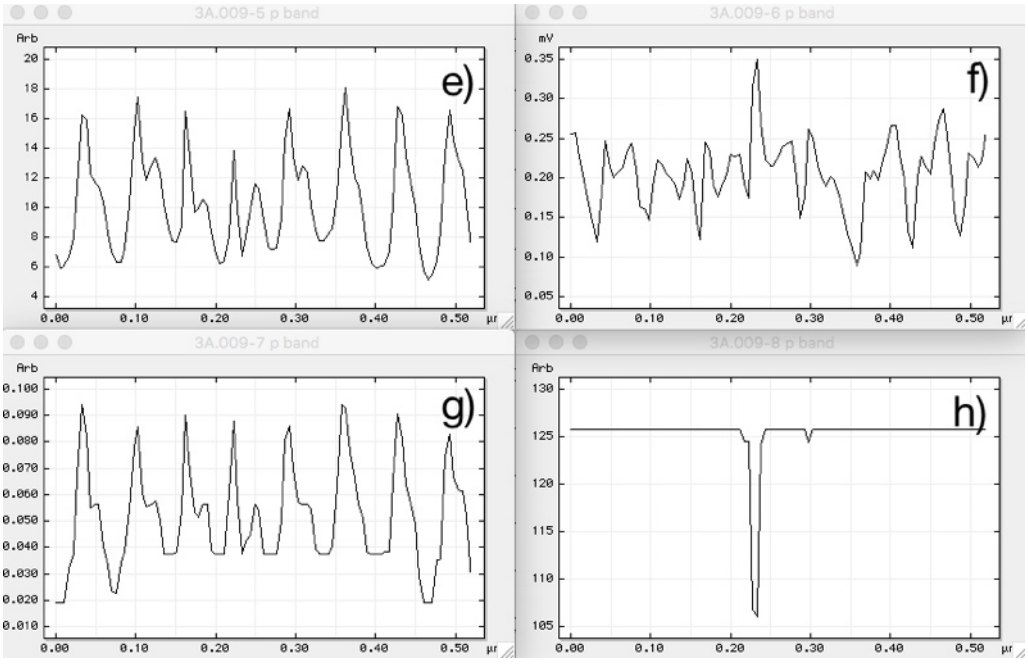


Figure 3.10: Plot profiles of channels 5 – 8 (e – h) in Fig.3.8. e) adhesion; f) deformation; g) dissipation; h) Sneddon modulus.

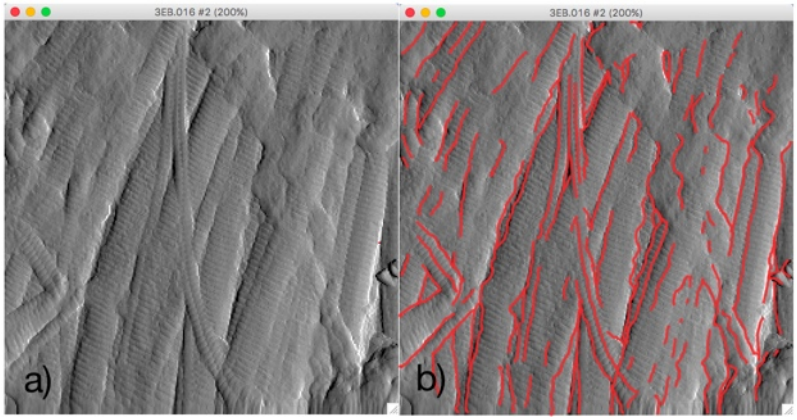


Figure 3.11: Typical AFM image of collagen-rich tissue. The fibre orientations vary, the fibres run above and beneath each other and the boundaries between them fade in and out. a) original scan; b) lines drawn over visible edges by hand. Note the degree of fragmentation of the lines.

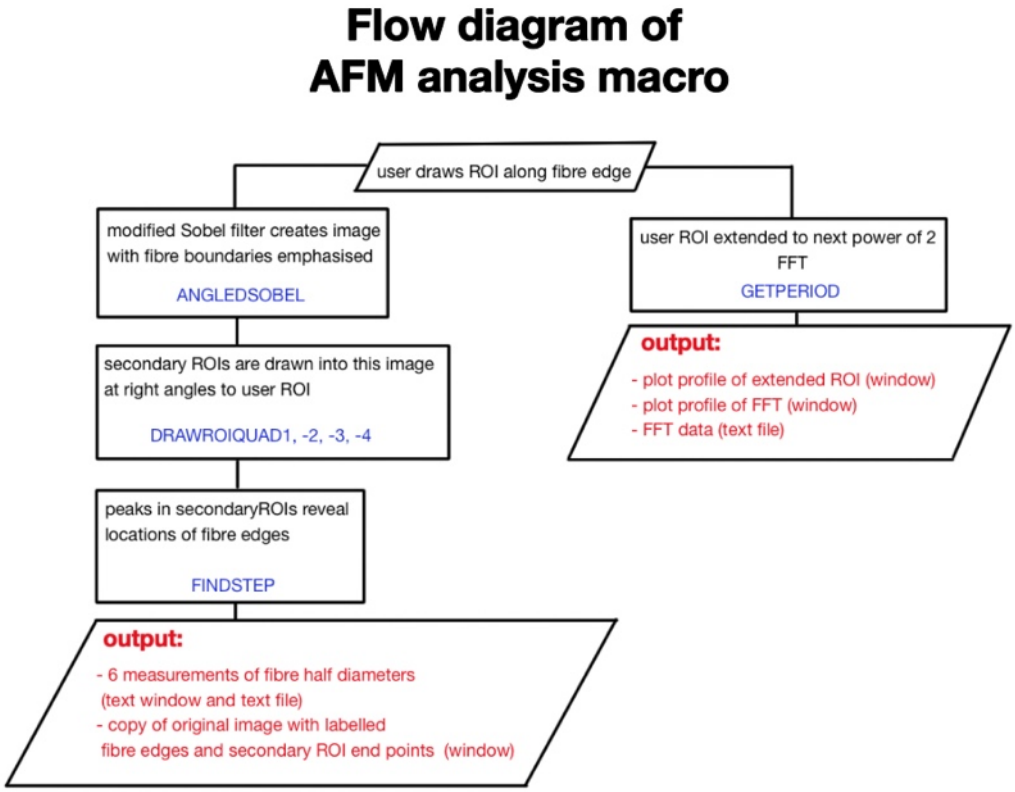


Figure 3.12: Brief outline of the macro algorithm. The names of the procedures at the relevant steps are written in blue. For clarity, some procedures have been omitted from the overview.

can be made within the copy image without affecting the original.) The ROI is copied into the copy image, measured and its number of length pixels is extended to the next higher power of 2. This is done in the array that stores the ROI's data points and allows a Fast Fourier Transform (FFT, see Appendix A.9) to be executed on the linear array. This FFT produces a line graph of frequencies that is stored as a data file, which can then be analysed by the user. Next, the macro measures the fibre thickness by extending secondary ROIs at right angles from the centre and both ends of the user-drawn ROI to the image edges. For this, a modified Sobel filter (see Appendix A.4) is applied to the copy image, which emphasises edges (i.e. steps in grey value in a perpendicular direction to the original user ROI) and de-emphasises grey value steps in the direction of the user ROI. See section 3.3.6. As a visual check that the ROIs have been done correctly, the macro draws red dots on the intersection points between these ROIs and the image edges. The plot profile of the ROIs are stored in an array and a procedure finds the pixel that first meets a condition for the identification of a fibre boundary. This condition is in principle a jump in grey value that can be set (*stepthresh*) over a certain number of pixels (*steplength*). When there is a great enough grey value difference within a short stretch of the ROI, it is assumed that the fibre boundary is reached. The appropriate values for these variables may differ between different image types. A further variable is the ROI width (*ROIwidth*), which when wider may make this signal clearer by averaging over a wider area. Green dots are drawn where the macro has identified such a fibre edge. The distance between the user ROI and the green dots on either side together constitute the fibre diameter at that position. For convenience, each of the two parts is termed "half", even though in reality the distance may diverge from exactly half of the diameter. When the condition isn't met anywhere along the ROI, the green dot is drawn at the origin of the secondary ROI, on top of the user ROI. A text window is created and saved containing the individual half-diameters, the diameters, the average diameter of the three measurements and the standard deviation.

For the macro code refer to the Appendix B.

3.3.2 Procedure *DrawRedDot*

A palette is loaded in which the first grey value is not white, but red. The remainder of the 255 values remains the usual gradient of greys to black. A small circular ROI is drawn and then filled with a grey value of '1' (i.e. red).

3.3.3 Procedure *DrawGreenDot*

A palette is loaded in which the second grey value is not nearly white, but green. The remainder of the other values remains the usual gradient of greys to black. A small circular ROI is drawn and then filled with a grey value of '2' (i.e. green).

3.3.4 Functions *degtorad* and *radtodeg*

These two functions convert degrees to rad, and vice versa. This is necessary, as the trigonometric functions work in rad, but the angles in Image SXM are returned in degrees.

To convert degrees into rad: $y = x * \pi/180$

To convert from rad into degrees: $x = y * 180/\pi$

3.3.5 Procedure *FindStep*

A secondary ROI perpendicular to the user ROI is drawn when this procedure is called. The image created by the *AngledSobel* procedure is used for this. A repeat...until loop is implemented, in which the ROI array is examined by moving along in one pixel steps. If the difference in grey value between the current pixel and the pixel at (current + *steplength*) and (current - *steplength*) stays below the threshold (*stepthresh*), nothing is done and the procedure moves on to the next pixel in the loop. If the grey value is above the threshold, the loop is exited and the pixel at which this occurred is set as the point at which a fibre edge is found and this value is given to a variable, *halfdiam* (one half of the presumed fibre diameter). User-defined parameters are *stepthresh* (the step in grey value above which a fibre boundary is assumed to have been found), *steplength* (the number of pixels over which a threshold is measured) and *ROIwidth* (the width of the secondary ROI). If this procedure cannot find a value that fits the criteria for an edge, *halfdiam* is given the value zero. For a closer look at these parameters see also 3.3.13.

3.3.6 Procedure *AngledSobel*

A Sobel filter detects edges by convolving the image with two kernels (see Appendix A.3) that emphasise grey value differences along the horizontal and the vertical axes and then combining them. For the purpose of this analysis the standard Sobel filter ([99]) was modified in such a way that it emphasises grey value gradients that occur at a certain angle in the image and de-emphasises grey value differences perpendicular to this angle. Removing the fibre banding from the image is desirable, as the grey value gradients associated with it can mimic fibre edges and lead to erroneous identification. This is done in the copy image and is achieved by calculating the magnitude of the grey value gradients in the x and y directions like in a standard Sobel filter, and then equating them to unit vectors in the x and y directions, mx and my, respectively. From these the magnitude of the sum vector m can be calculated via Pythagoras' theorem $m^2 = (m_x)^2 + (m_y)^2$ and its angular direction θ determined via the arctan function. If the angle along which the edges of interest are suspected to run is α, then the magnitude of the sum vector from each kernel can be multiplied by $\sin(\alpha - \theta)$, which is 1 when θ is at right angle to α, and 0 when $\theta = \alpha$. Hence, when the overall gradient is in line with the user ROI (and the fibre long axis) it gets de-emphasised by multiplication with a low value, and when it is close to perpendicular to the user ROI (and the fibre axis), it gets emphasised, relatively, by multiplication with a value close to 1. Grey value steps that are encountered at the fibre edges with a gradient close to $90^o + \alpha$ are therefore kept prominent, whereas the banding

that shows grey value gradients at angles close to the fibre axis disappears. See fig. 3.13 and fig 3.14.

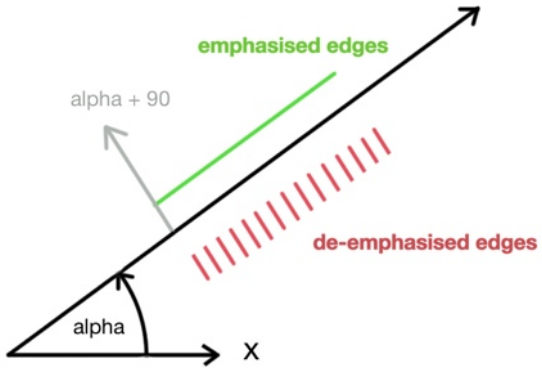


Figure 3.13: Illustration of how edges are emphasised and de-emphasised by the angle-constrained Sobel filter, depending on the measured angle of the user ROI.

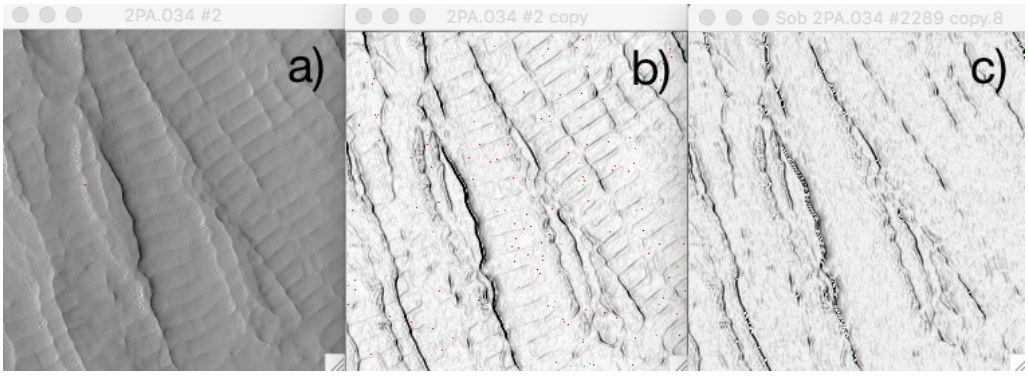


Figure 3.14: Comparison between the original image (a), the conventional (b) and the angle-constrained Sobel filter (c). Note the absence of banding in c.

3.3.7 Procedures *DrawROIquad1, -2, -3, -4*

These four procedures draw the secondary ROIs that are perpendicular to the user-drawn ROI and serve to find the edges of the fibres. The difference between *DrawROIquad1, -2, -3* and *-4* is into which quadrant the initial user ROI is drawn. If a user, for instance, draws a ROI from the bottom left into the top right corner, this is regarded as quadrant 1, and the secondary ROI will be drawn into quadrant 2, but if the user draws from the top right into the bottom left corner, this would be regarded as quadrant 3, and the secondary ROI would be drawn into quadrant 4. (Note that the quadrants refer to the angles within a circle, not the location of the ROI. The horizontal corresponds to an angle of 0 degrees, and positive angles are measured from there anticlockwise. A ROI drawn, for example, into quadrant 4 can be anywhere in the image, but its measured angle will always be between 270 and 360 degrees.) For reference see fig. 3.15.

For ease of understanding and, if necessary, debugging, these situations were separated, rather than trying to incorporate all situations into one procedure. As Image SXM returns

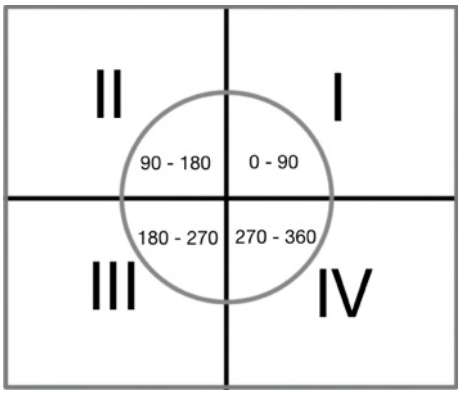


Figure 3.15: Correspondence of angles with quadrants in a unit circle.

angles from 0 to 360 degrees, depending on where the starting and end points of the ROIs are, these angles were maintained and not converted to 0 -180 or -90 - +90 degrees. One function of the *DrawROIquad* procedures is to find which side of the image edge the secondary ROI intersects. This is important in finding the end points of the secondary ROIs. See fig. 3.16 for an example where the user ROI has been drawn into quadrant I. This means that the first secondary ROI will be drawn into quadrant II. h1 and h2 are the hypotenuses of the two right-angled triangles of which the y and x positions of the starting points (on the user ROI) form the cosines. Which of these hypotenuses is larger distinguishes whether the end point of the secondary ROI intersects the image edges on the horizontal or the vertical. The secondary ROIs are always drawn anticlockwise from the user ROI. Thus, a user ROI drawn into quadrant I (0 - 90 degrees) will be followed by a secondary ROI drawn into quadrant II (90 -180 degrees), a user ROI drawn into quadrant II will be followed by a secondary ROI into quadrant III, etc. The *DrawROIquad* procedures first find the end points of the secondary ROIs, then the red dot indicating the end point is drawn into the copy image (for control purposes). A ROI is then drawn from the start to the end point and measured. This measurement fills the plot data array with the grey values of the pixels along the ROI and also returns the angle of the ROI. Then the procedure *FindStep* is called (see above) and the ROI is killed in preparation for the next ROI, which is drawn into the same quadrant further along the user ROI. When all secondary ROIs have been drawn into this quadrant, the three secondary ROIs are drawn into the opposite quadrant. Finally, with the value returned from the *FindStep* procedure, the position of the presumed fibre edge along this ROI is calculated and a green dot drawn there, by calling the *DrawGreenDot* procedure.

3.3.8 Procedure *GetPeriod*

The *GetPeriod* procedure is called early on from the main procedure *AFMmeasure*. It uses the data from the original user ROI to find the periodicity of the banding of the fibre along which the user ROI has been drawn. The periodicity is calculated via a one-dimensional FFT, however, this requires the data to be a length of a power of 2. As user ROIs will generally not be exactly this length, the data array in which the ROI grey values are held is extended to the next higher power of 2, and the additional array positions are filled

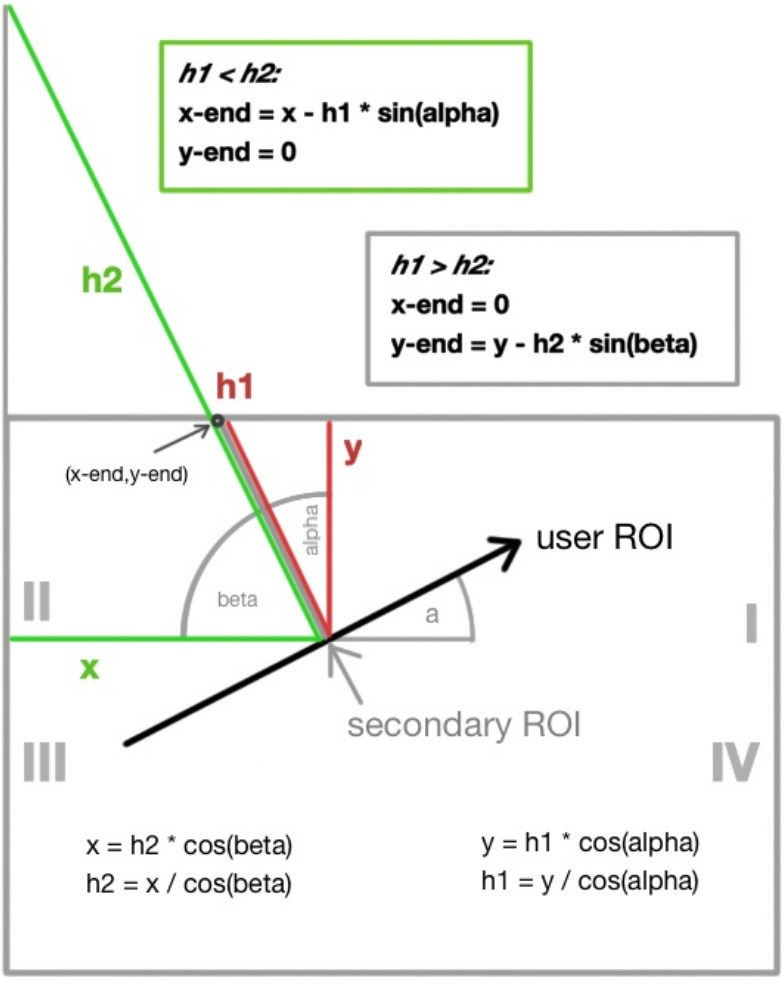


Figure 3.16: Illustration of how the end points of the secondary ROIs are derived from the user ROI's angle a.

with a mirror image of the original data. Various options for filling the extension of the array were considered, among them re-using the value of the last original pixel or using a mid value of 127, but the creation of a sharp feature at the extension point is likely to produce artifacts in the FFT (sharp edges lead to many different frequencies), and so the least intrusive extension, which would furthermore retain some of the original frequency elements, was deemed to be a mirroring of the original data around the end point of the original ROI data. For practical reasons (limited access to inbuilt arrays in the macro language), these new data are then transferred into the linebuffer array and inserted into a new image, in which a 2^n ROI can be drawn and a one-dimensional FFT executed on its new plot profile. This plot profile, and the plot profile of the FFT are displayed in new windows, and the data of the FFT exported to a text file and saved. For an example, see fig. 3.17. Due to the fact that the macro language has no direct access to the FFT's plot profile array, the user has to identify the main frequency via the plot profile window or the saved data file.

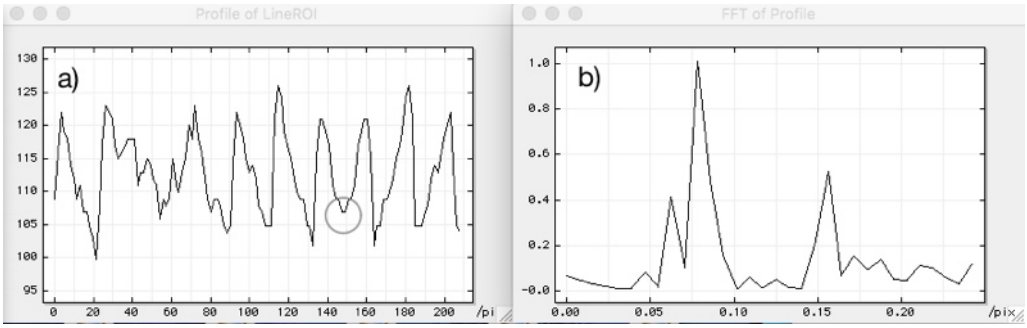


Figure 3.17: Examples of the extended ROI's plot profile and the plot profile of the one-dimensional FFT of this. The grey circle indicates the mirror point, where the drawn line ROI's plot profile ends and is extended by its mirror image. The main peak in the FFT at around 0.076/pixel corresponds to a length scale of about 13 pixels, which is not easily recognised in the line profile of the ROI. The smaller peak at around 0.155 corresponds to about 7 pixels and is likely a harmonic of the first. The obvious periodicity of around 20 pixels in the line ROI profile corresponds to the small peak just under 0.05/pixel in the FFT plot.

3.3.9 Procedure *DrawMainROI*

This procedure draws a copy of the user ROI as a black line into the copy image in order to help the user remember where they drew their ROI in relation to the rest of the image and the fibre edges detected.

3.3.10 Procedure *DrawAllROIs*

This procedure governs which *DrawROIquad* procedure is called when. In each analysis the *DrawROIquad* procedures are called six times, one for each secondary ROI drawn. The sequence is illustrated in fig. 3.18, where the black arrow represents the user ROI, the red dots the starting points of the secondary ROIs and the grey lines the secondary ROIs. The numbers correspond to the sequence in which they are created by the macro. The

first secondary ROI starts in the centre of the user ROI and is drawn from there to the image edge into the quadrant that follows that of the user ROI, then the second secondary ROI is drawn from the same starting point into the opposite quadrant, to the image edge. Both form a line perpendicular to the user ROI through its centre. The following ROIs, 2a and 2b, are drawn in the same way but starting at the origin of the user ROI, and ROIs 3a and 3b are drawn with the starting point being the end of the user ROI.

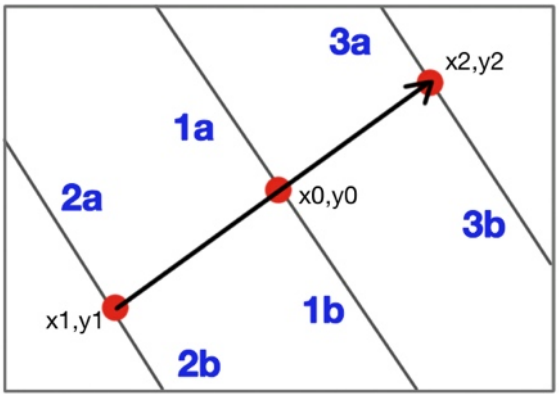


Figure 3.18: Sequence in which the macro draws the secondary ROIs. The black arrow represents the user ROI, the grey lines the secondary ROIs and numbers indicate the sequence.

3.3.11 Procedure *AFMmeasure*

This procedure is the heart of the analysis. It sets up all necessary housekeeping operations, such as ensuring that the y axis is not inverted (by default Image SXM inverts the y axis so that 0 is at the bottom of the image), that the x,y scale of the image is in pixels and that the z scale is in grey values ranging from 0 to 255. It checks if a line ROI has been drawn, gives the windows an ID number, creates copies of images and names them appropriately, etc.

At first, a copy is made of the original image, the user ROI is restored into it and measured. This yields the x,y coordinates of the ROI centre, its length and angle. *GetLine* furthermore returns the starting and end points of the ROI and its width.

Next, the *GetPeriod* procedure is called (see above), which extends the ROI data to the next higher 2^n, produces a plot profile of this new ROI and of the FFT of this, and saves the FFT's plot data in a text file.

After this, the angle of the user ROI, a, is converted into the angle α (which depends on the quadrant a is in) and the *AngledSobel* procedure (see above) is called. In this, the copy of the original image is subjected to an angle-constrained Sobel filter, through which the fibre edges should be emphasised, and the banding edges de-emphasised. Into this image a line in the position of the original user ROI is drawn by the *DrawMainROI* procedure (see above).

In the next step, *DrawAllROIs* calls the *DrawROIquad* procedures one by one in the right sequence, which draw all red and green dots into the image and returns values

for the distance between the user ROI and the presumed fibre boundary (half diameters called *halfdiam*). These *halfdiam* values are then written into a text window, added to their corresponding half diameter to give three diameters for each centre, starting and end position on the user ROI. The average diameter is also calculated, as is the standard deviation. This text window is then saved.

3.3.12 Macro *AFMmeasureWithLoop*

Although the *AFMmeasure* procedure does the main work, the user-callable macro portion of the code is this section, which only serves to run the above algorithm a number of times with different parameters. The parameter *stepthresh* is aided by a parameter *threshst* that allows *stepthresh* to be run in steps > 1, since the grey value differences across fibre boundaries after the *AngledSobel* procedure can be large. The parameter *steplength* allows to define the length over which *stepthresh* is measured, and *ROIwidth* is an additional option to set the ROI width used to find the edges to values > 1. To allow larger steps also for *steplength* and *ROIwidth*, their equivalent of *threshst* are *lengthst* and *ROIwidthst*, respectively. For illustration, fig 3.19 shows the output of the macro. For explanation see the figure legend.

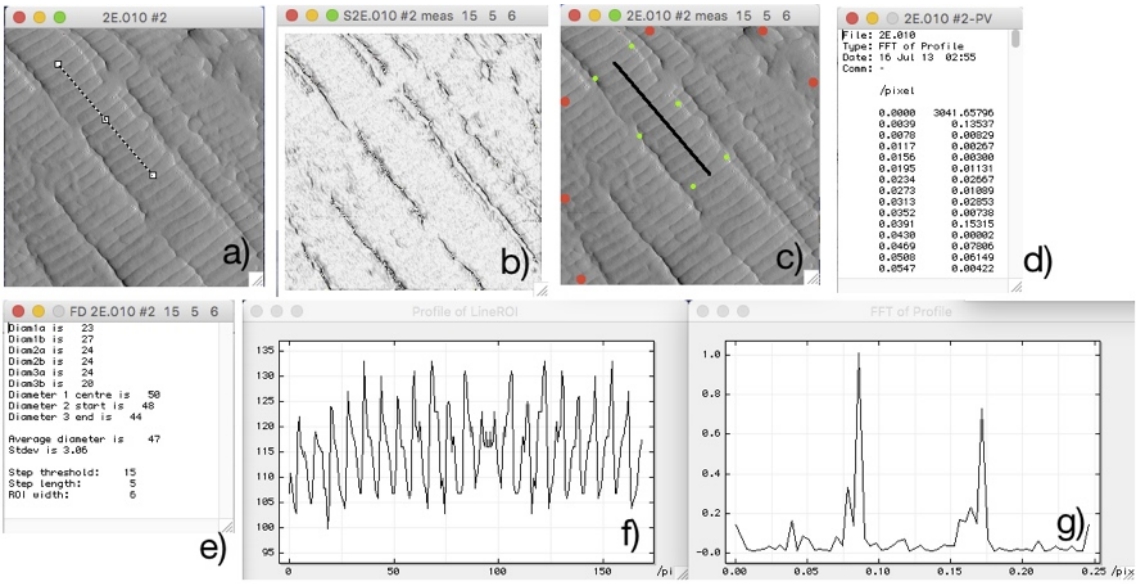


Figure 3.19: The output of the AFM measurements macro: a) original image with the user ROI drawn in; b) angle-constrained Sobel filter applied to image shown in a); c) copy of original image with green dots indicating found fibre edges and red dots to show where the ROIs have been drawn; d) text file with Fourier frequencies; e) text file with half diameters, diameters, average and standard deviation; f) plot profile of line ROI extended to next 2^n; g) frequency plot of data from d).

3.3.13 Analysis variables

The user-definable parameters, *stepthresh*, *steplength* and *ROIwidth*, are all related to the finding of fibre edges in the *FindStep* procedure.

The *stepthresh* parameter: the algorithm of *FindStep* is based on the observation that, after the angle-constrained Sobel filter has been applied, the nearest fibre edge is identified by a marked spike in grey values. See fig. 3.20.

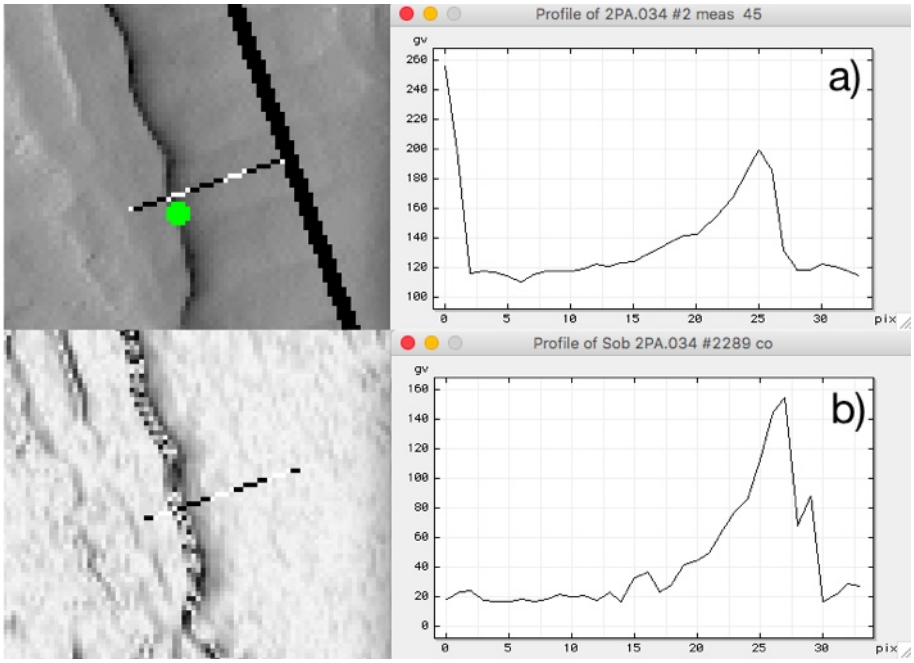


Figure 3.20: After the angle-constrained Sobel filter has been applied, a larger, more pronounced peak identifies the fibre edge (b), compared to the unfiltered image (a).

The *stepthresh* parameter, short for step threshold, is the difference in grey value at a position along the ROI and that at a position a defined distance (*steplength*) away from it. If a step threshold is chosen at the right size for the peak - over the chosen distance of the following parameter, *steplength* - then the macro recognises the first instance this peak occurs as indicative of the fibre edge.

The *steplength* parameter: this parameter is essential for the relevance of the previous parameter, *stepthresh*, as it defines the narrowness the peak needs to have in order to be recognised as significant of a fibre edge. A grey value difference that occurs gradually over a long distance is not likely to be an edge, and likewise a sharp peak may be missed if the step length is so long that it spans from a peak to a region too far from it. If the step length is too short, no significant difference in grey values may occur, even though the peak is large, when it is not sufficiently sharp. This is illustrated in fig. 3.21.

The *ROIwidth* parameter: a ROI of one pixel width consists at a certain position along its length only of the value of the pixel at that position. A ROI of width n at the same position has the value of the mean of those n pixels at that length position. This can be an advantage or a disadvantage (see below). It is an advantage in cases, for instance, where the edge is not continuous, but fragmented or with varying grey values. In this situation, as it cannot be foreseen where exactly a secondary ROI meets a fibre edge, it could be advantageous to draw a thicker ROI in order to ensure that a wider area of the fibre boundary is sampled.

As the nature of images and samples may differ considerably, it is advisable to run

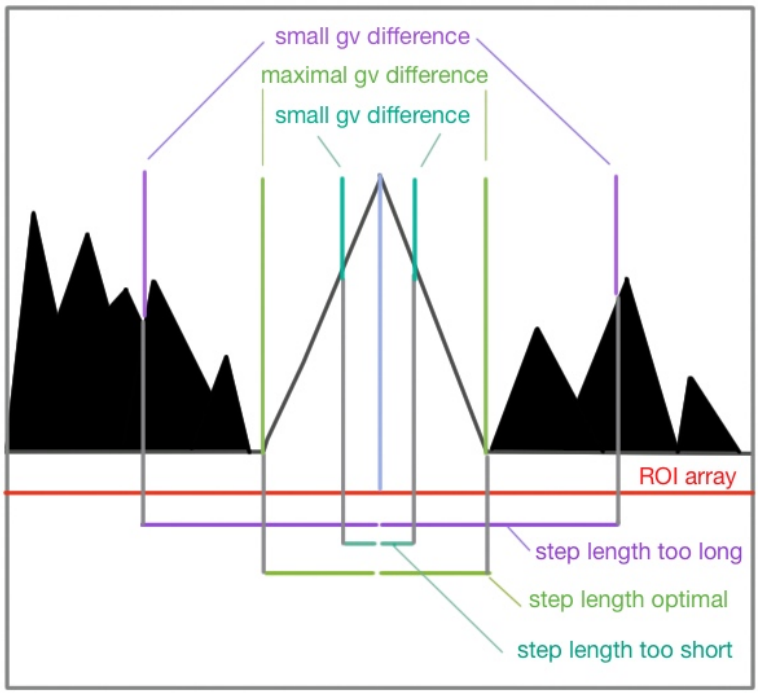


Figure 3.21: Finding the right value for *steplength* is crucial for maximising the grey value difference detected from a given peak. The step length shown in green detects the maximum grey value difference, whereas the blue and purple step lengths are too small and too large, respectively, to find substantial differences.

through a reasonable number of parameters for each new sample or image type to determine appropriate parameter values. Provision for this is made in the main macro.

In theory, *ROI width* could also be varied for the analysis of banding periodicity in *GetPeriod*. However, as the banding stripes do not always run at right angles to the long fibre axis, and a line ROI can not always be drawn exactly perpendicular to the lines of the fibre bands, a line ROI wider than one pixel would be detrimental in such situations. In a line ROI that is, for example, 30 pixels wide, the plot profile for every pixel of length is made up of the grey values of the 30 pixels at this length position. If the banding is not perfectly perpendicular to the ROI direction, each position on the plot profile of a 30 pixel wide ROI is made up of grey values that belong to different positions on the wave produced by the banding. Thus, several phases contribute to the same position, resulting in a deterioration of the signal. For example see fig 3.22.

3.4 Evaluating the macro

3.4.1 Frequency measurements

The macro measures the frequency components of the line ROI the user has drawn, extended to the next higher 2^n, via the inbuilt FFT function. The output is both a plot profile window that Image SXM creates automatically, and a text file containing the data values of the plot. The user has to analyse these two outputs manually and select the rel-

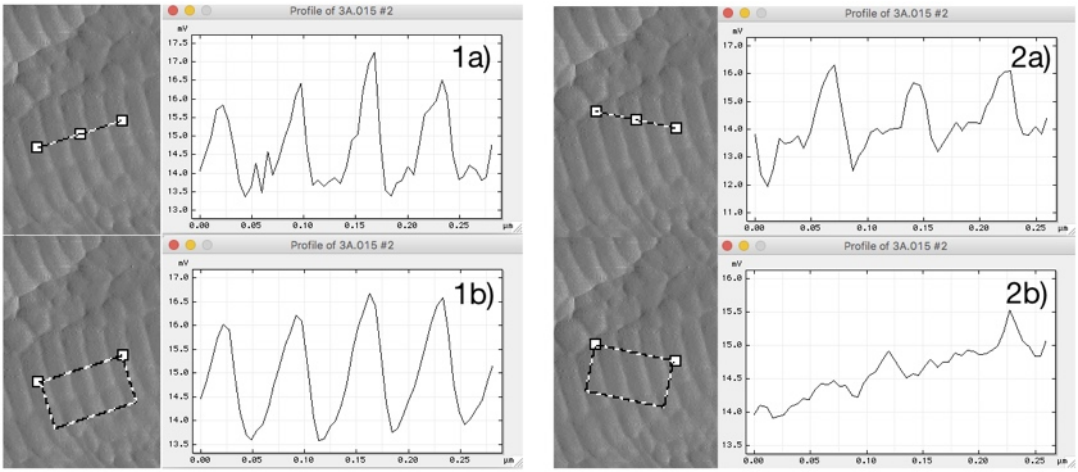


Figure 3.22: These two examples show how a wide ROI can positively or adversely affect how clearly periodicity shows in a plot profile. In 1a) the ROI is drawn perpendicular to the banding, which is clearly visible in the plot profile. In 1b) the same ROI is widened to 30 pixels. The banding is even more obvious. In 2a) the ROI is drawn at an angle across the same fibre, which changes the plot profile somewhat, but still shows some periodicity. In 2b) the same ROI as in 2a) is widened to 30 pixels, and the periodicity in the plot profile is lost.

evant peak in the FFT. The typical banding periodicity of collagen type I is 67nm, which corresponds to 11.5 pixels in these 256 pixel (1.484μm) square images, when measured parallel to the sides. 11.5 pixels correspond to 0.087 /pixel on the FFT's frequency-axis. A peak at this position is indeed present in all images. A selection is shown in fig. 3.23). Although this peak is present, it is not always the only, and at times not even the main peak in the FFT. Other peaks may be due to harmonics of the same frequency, or to other elements in the image, such as for example scan resonances. Within the collagen banding there is also substructure that will give rise to other frequencies ([100]).

3.4.2 Fibre diameter measurements

Every run of the macro produces 6 distance measurements, each corresponding to the distance between a point on the user ROI and the fibre's edge. These individual measurements are designated as 1a for the distance between the ROI centre and the edge, anticlockwise from the quadrant into which the user ROI has been drawn, 1b from the same point clockwise, 2a from the start of the user ROI to the fibre edge anticlockwise, 2b from the start clockwise, 3a from the end of the user ROI to the fibre edge anticlockwise, and 3b from the end clockwise - see fig. 3.18. As these represent individual measurements, for testing purposes, they are not combined into three whole diameters, but investigated separately.

The three parameters, *stepthresh*, *steplength* and *ROIwidth*, are likely to have different optimal values, depending on image type. Predicting what these values should be is not trivial, and so to approach any optima, it is necessary to explore the parameter space by running the macro through a large number of parameter combinations.

In order to get a first hint at which parameter sets might yield usable results, one peak

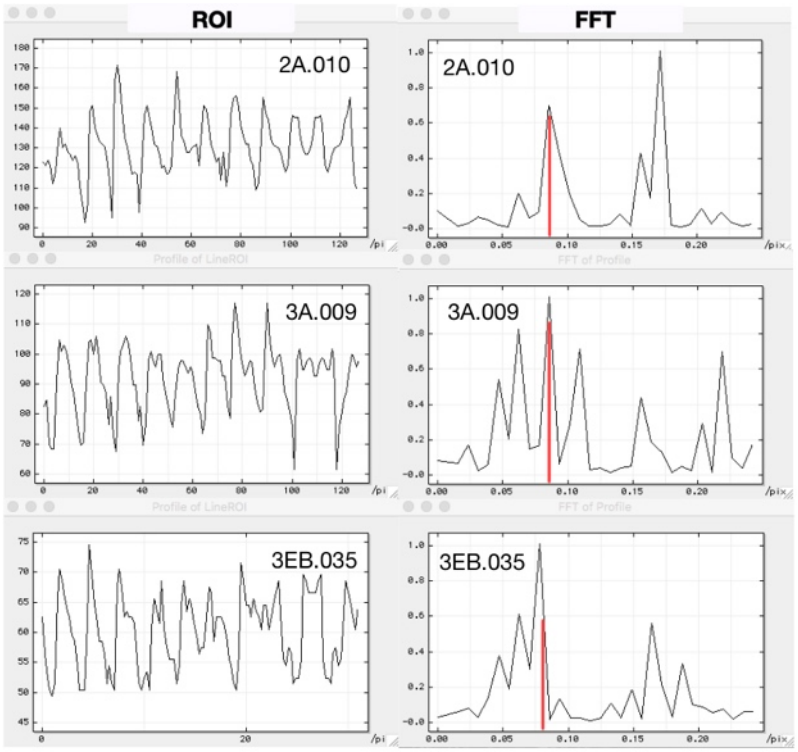


Figure 3.23: Some examples of ROI plot profiles and corresponding FFT plot profiles from macro output. The ROI plot profiles are in the left column, the corresponding FFT plot profiles in the right column. The frequency corresponding to collagen banding is highlighted in red. The slight shift away from this frequency in the last example may be due to the ROI being drawn at a slight angle to the fibre axis.

force error image was chosen initially and the macro run only on this single image through a wider parameter range than expected to be useful. *stepthresh* was taken from 5 to 100 in steps of 5, *steplength* from 2 to 20, in steps of 1 between 1 and 8 and then in steps of 2 between 10 and 20, and *ROIwidth* from 1 to 20 in steps of 1. The separate measurements 1a, 1b,2a, 2b,3a and 3b were reproduced by hand and the differences between macro results and these manual results noted.

Although manual measurements are not truly objective, they are as good as current measurement options can get, as the human visual system is still much superior to machine algorithms in taking in the patterns in the whole image. Three measurements were made by hand for each of the 6 positions (1a, 1b,2a, 2b,3a and 3b), by drawing a line ROI and reading off its length, at a magnification of 400%. This was done for the single image, in which a wider range of parameters were tried, as well as for each of the 10 images, which served to test the macro. The two main sources of error in these measurements were manual inaccuracies in the placement of the cursor, and real uncertainty as to where the true fibre boundaries lay. The spread of these three measurements were rarely more than 1 pixel.

Where the macro was unable to find an edge along a secondary ROI, a value of zero was assigned for the "half-diameter". Thus, zero is used as a flag, rather than a measured result. The number of measured, non-zero values returned by the macro in each run was called DifN. This value can range from 0 (when no edge is found) to 6 (when an edge is found in all six ROIs). For all non-zero values the difference to the manual measurement was calculated and then averaged over the run. This average DifAvg can be constituted of up to 6 measurements (DifN = 6). Only DifAvg values made up of 6 measurements were used to grade the success of the macro. Likewise, DifAvg values of no higher than 1.5 pixels were included in the following assessments. The first results of the wider parameter range run on a single image (2PA-034-2) are shown in figs 3.24, 3.25 and 3.26.

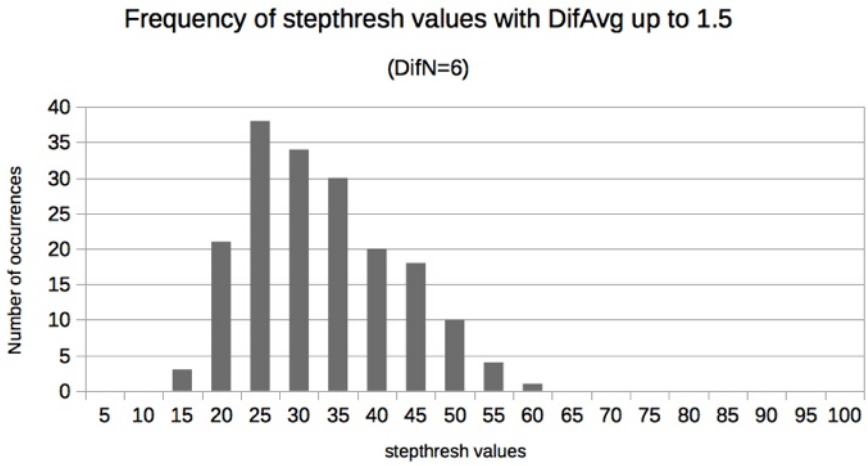


Figure 3.24: Histogram of *stepthresh* grey values resulting in DifAvg values below 1.5 with DifN=6. Data from one image (2PA-034-2). n=180.

Thus, sensible *stepthresh* values range up to about 50, *steplength* values from 3 to

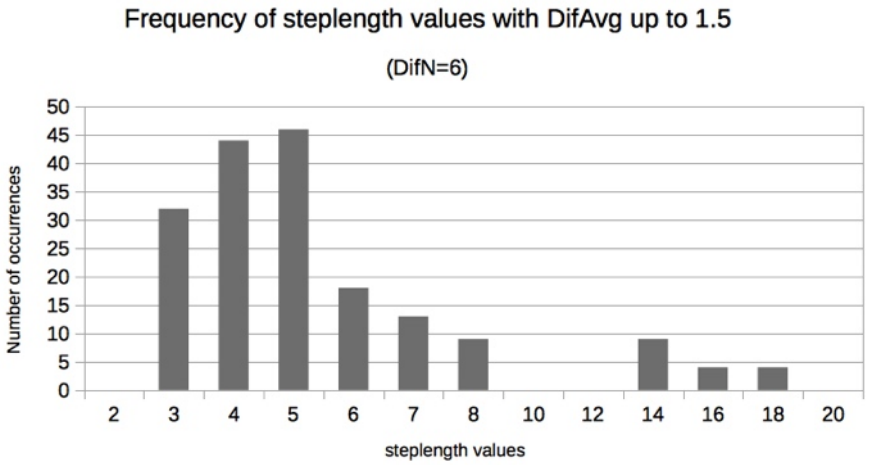


Figure 3.25: Histogram of *steplength* pixel values resulting in DifAvg values below 1.5 with DifN=6. Data from one image (2PA-034-2). n=180.

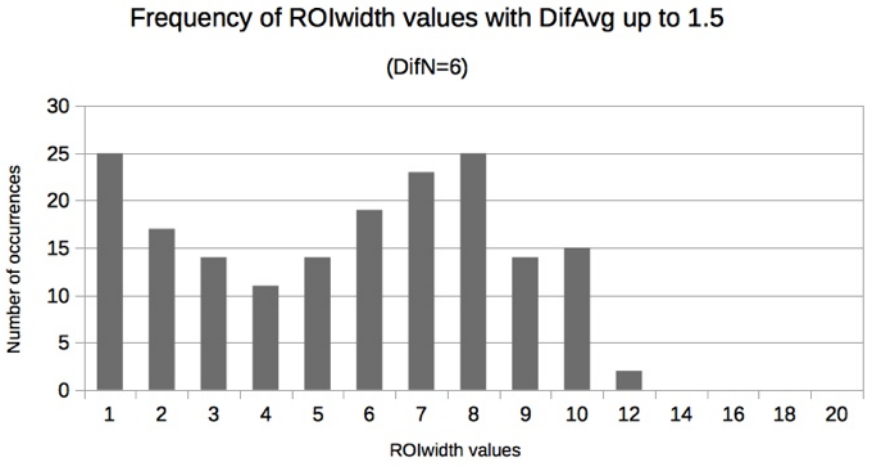


Figure 3.26: Histogram of *ROIwidth* pixel values resulting in DifAvg values below 1.5 with DifN=6. Data from one image (2PA-034-2). n=180.

about 8, and *ROIwidth* values from 1 to 10.

All 5 samples with both height and peak force error channels (10 images) were chosen to investigate the results using this new parameter range. Following the initial runs, *stepthresh* was now taken from 5 to 50 in steps of 5, steplength from 3 to 8, and *ROIwidth* from 1 to 10. This resulted in 600 runs for each of the 10 images.

As a value of zero is given when no edge could be found, as a first criterion for the general usefulness of the macro the number of zeros returned in each image from all parameter sets, was used. The numbers are shown in fig. 3.27. Here, N indicates the number of zeros returned, and the entries the number of runs where N zeros were returned. Hence, the more entries fall into the category where N has a low number, i.e. where the macro rarely fails to detect an edge, the higher the success of the macro. The possibility of the macro falsely detecting an edge is eliminated by only including the measurements where the average difference between manual measurements and macro measurements, DifAvg, is smaller than 1.5 pixels.

File	N=0	N=1	N=2	N=3	N=4	N=5	N=6
2A-010-1	318	93	87	59	16	27	0
2A-010-2	328	98	64	80	26	3	1
2E-010-1	275	65	91	52	36	72	9
2E-010-2	590	8	2	0	0	0	0
2PA-034-1	338	71	35	62	83	11	0
2PA-034-2	597	3	0	0	0	0	0
3A-009-1	143	109	56	31	57	46	158
3A-009-2	590	10	0	0	0	0	0
3EB-035-1	437	94	54	15	0	0	0
3EB-035-2	550	50	0	0	0	0	0

Figure 3.27: Success rate of the macro over all parameter sets. Column of N=0: the number of parameter sets that found 6 edges and thus returned no zero; column of N=1: the number of parameter sets that found 5 edges and returned 1 zero; and so on until column N=6: number of parameter sets that found no edges and returned 6 zeros. The weighting is strongly towards low N numbers, indicating a good success rate. Only measurements with DifAvg ¡ 1.5 pixels are included.

The high number of parameter sets in the N=0 column is a good first sign, indicating that the macro algorithm is capable of finding edges under a variety of different parameter conditions, and that a return of zero is a minority result. However, it is also evident that there is a wide gap between the results for different images. Although 3A-009-1 and 3A-009-2 are two channels of the same sample, their number of found edges diverges widely. Overall, there appears to be a tendency for peak force error channels (file names ending in -2) to have a higher success rate. But even comparing height and peak force error images within their own group, there is still divergence. In practice, a large factor in determining the success of the edge finding lies in the choice of fibre, and in the location along the fibre, in which the user ROI is drawn (compare fig. 3.11). This may account for some of this divergence.

When comparing the results for the different images with each other, the value of DifAvg was kept as pixels, and not a percentage, since firstly the half diameter lengths were very similar in all images, and secondly the absolute lengths in this case have little

meaning in terms of the error, which occurs mainly at the location of the fibre edge and this environment is very similar, irrespective of the distance between it and the user ROI.

As an overview over the results of all images and parameters, figure 3.28 a) shows the distribution of the DifAvg values between 0 and 100 that are constituted of 6 measurements (DifN=0 to DifN=5 not included.). Fig. 3.28 b) shows a section of a) for DifAvg values between 0 and 2 in greater detail. The total number of values included in the latter is 4977.

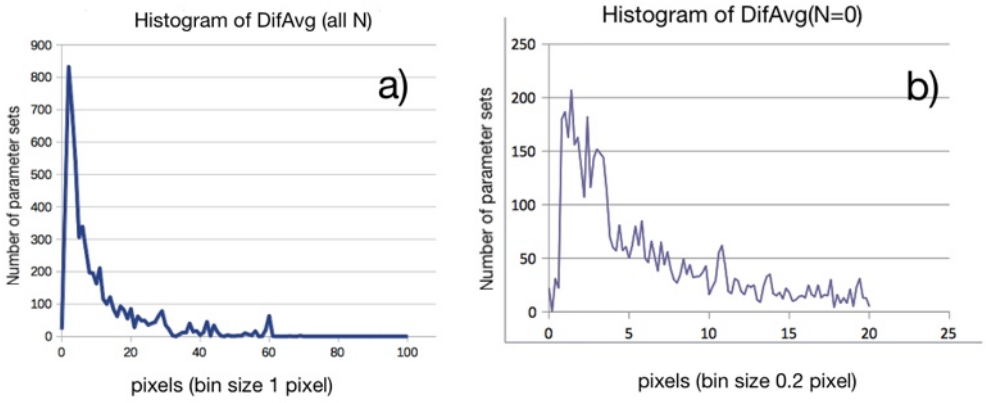


Figure 3.28: Histogram of the DifAvg values, a) from 0 to 100 pixels. Bin size is 1 pixel. b) from 0 to 2 pixels. Bin size is 0.2 pixels. n=4977. Note the significant peak within the first five bins in b), corresponding to a < 1 pixel difference between manual and macro measurements.

Like fig.3.27, it is again reassuring that the values clearly peak at very low DifAvg values, indicating that the majority of the macro results closely resemble the manual measurements.

Similar to the figures relating to the initial measurement of one image, figs 3.29, 3.30 and 3.31 show the new parameter set run on all images. Results with DifAvg up to 1.5 and DifN=6 are shown.

Only the histogram for *stepthresh* (fig. 3.29) shows a peak, indicating that some *stepthresh* values might be more successful than others. The histograms for both *steplength* (fig. 3.30) and *ROIwidth* (fig. 3.31) show no noticeable preference for any particular parameter value, and only *steplength* shows a weak trend towards lower values.

As these data for one parameter combine the other two parameters in them, it is conceivable that particular, successful combinations might hide in the data. In order to see whether within the three highest steptresh values (30, 35 and 40 pixels) other parameter values have any predominance, the frequencies of *steplength* (fig. 3.32) and of *ROIwidth* (fig. 3.33) were plotted.

No substantial predominance of any secondary parameter is seen in these *stepthresh* values. The same is true for the other parameter combinations (not shown). This indicates that the macro does not seem to be very sensitive to parameter values, suggesting it might be able to deal successfully with a wide range of images.

Although the success rate is high, there is still a non-negligible number of cases where

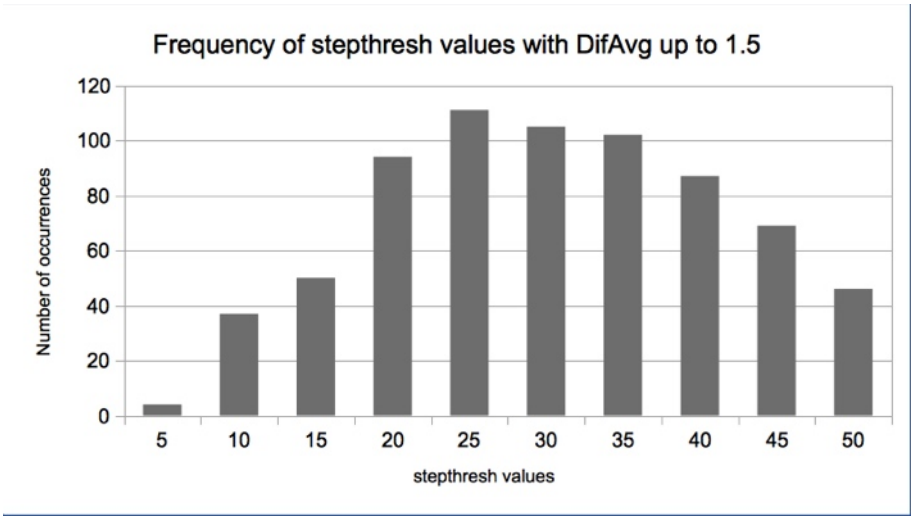


Figure 3.29: Histogram of *stepthresh* values resulting in DifAvg values below 1.5 with DifN=6. All images and all values for *steplength* and *ROIwidth* are included in the data. n=706.

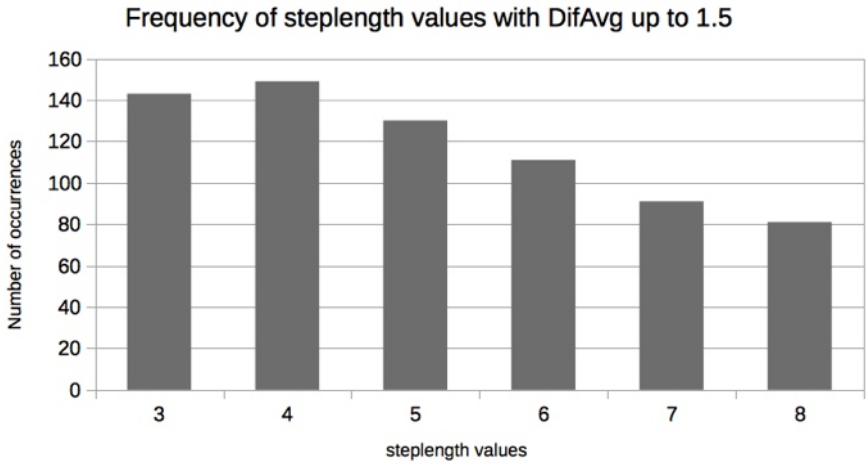


Figure 3.30: Histogram of *steplength* values resulting in DifAvg values below 1.5 with DifN=6. All images and all values for *stepthresh* and *ROIwidth* are included in the data. n=706.

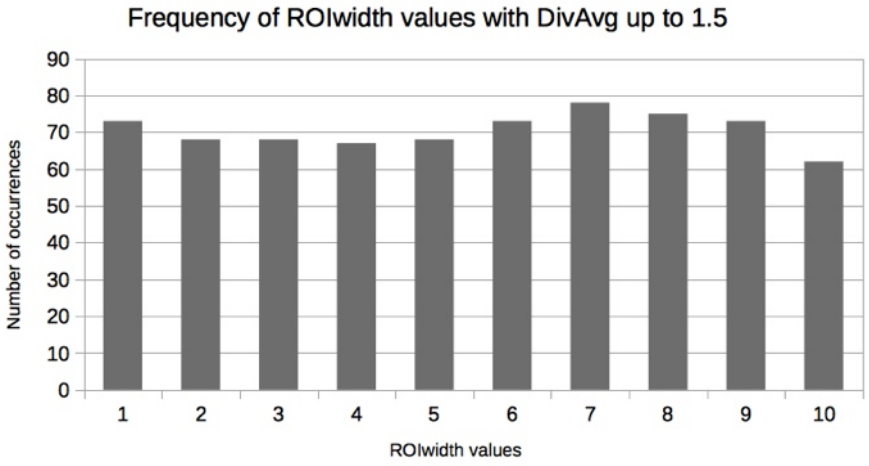


Figure 3.31: Histogram of *ROIwidth* values resulting in DifAvg values below 1.5 with DifN=6. All images and all values for *stepthresh* and *steplength* are included in the data. n=706.

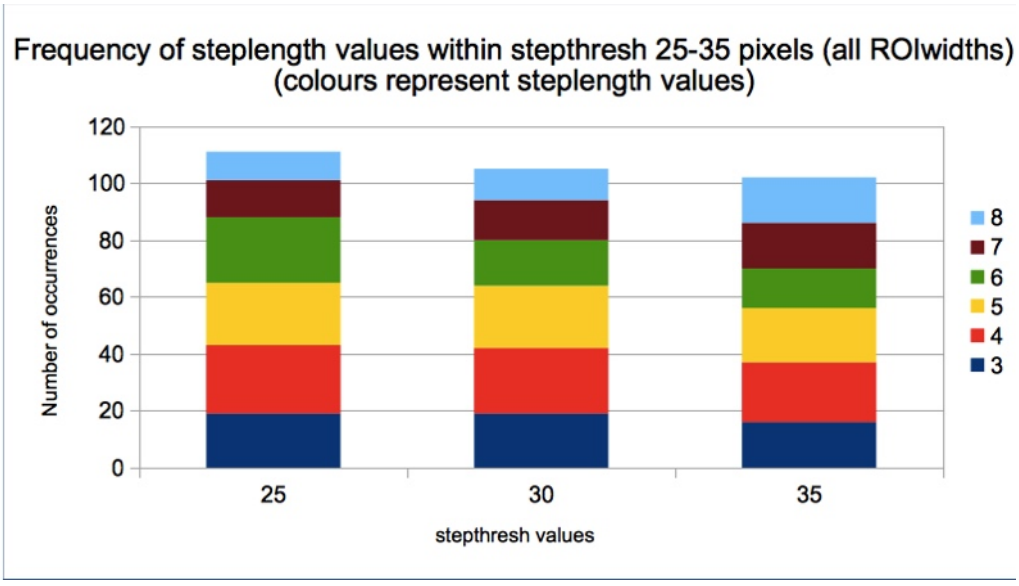


Figure 3.32: At the most successful step lengths, 25, 30 and 35 pixels, the *steplength*s producing DifAvg values below 1.5 appear to be relatively uniformly represented.

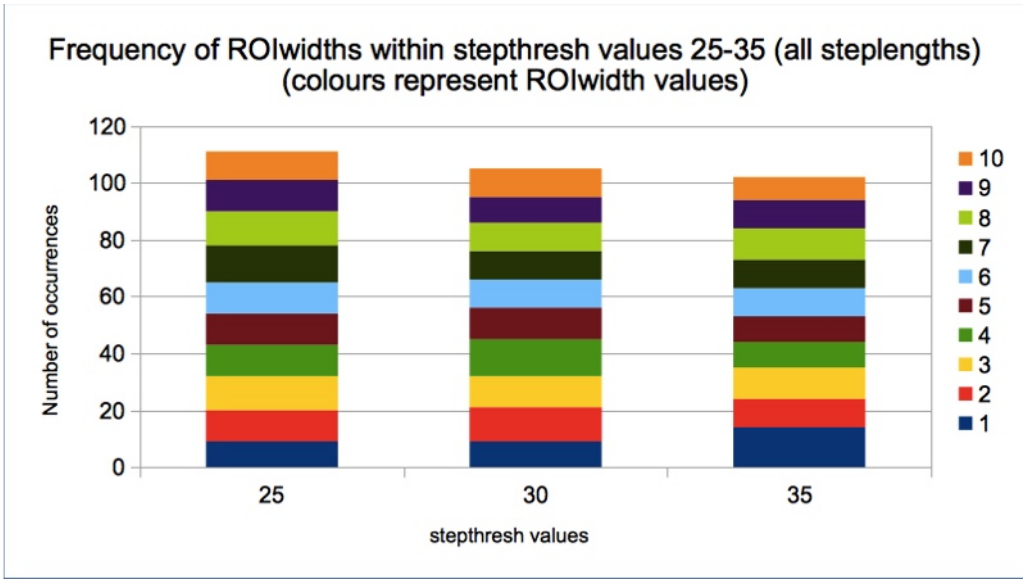


Figure 3.33: At the most successful step lengths, 25, 30 and 35 pixels, the *ROIwidth*s producing DifAvg values below 1.5 also appear to be relatively uniformly represented.

the macro does not find any edge, or identifies a wrong edge. This appears mostly due to where the ROI is drawn by the user, as some fibre edges do not show sufficient contrast along some stretches of their length.

To illustrate this sensitivity, two fibres were measured in the following way: a ROI was drawn along the middle of the fibre and the macro run. Then the same ROI was elongated slightly and analysed again. This was repeated 12 times. The starting position of the ROI was kept constant. As the ROI's middle and end move to different positions along the fibre, the macro, using identical parameters, succeeds or fails in finding edge points depending on where along the fibre it measures. Fig. 3.34 shows a fibre that has a clear and sharp edge along most of its length. Fig. 3.35 shows a fibre in the same image whose edge features are more eroded. The green dots indicate the points detected as edges by the macro. The parameters used were *stepthresh* 30 pixels, *steplength* 4 pixels and *ROIwidth* 3 pixels. It can be seen from these images, and from fig. 3.36 that the success rate varies along the fibre, depending on the local environment at the point where the secondary ROIs measure the grey value steps.

This clearly shows that the only direct human input in the macro is an important source of variability in its success rate. Although the user may become more adept at this, there is often a limit to the choice available in the image, so this is not an ideal situation. Ways to alleviate this problem will be discussed in section 3.5.

In spite of its sensitivity to the choice of ROI, the algorithm is nevertheless promising. It is not only largely insensitive to variations in parameters, which will allow the macro to deal successfully with a number of image types. But even at the present state, the output of individual half diameter measurements in combination with the green dots that allow instantaneous judgement of failure or success makes it possible for the user to decide which measurements to accept and which to reject.

At the present stage, the macro is not yet suitable for a high throughput of images,

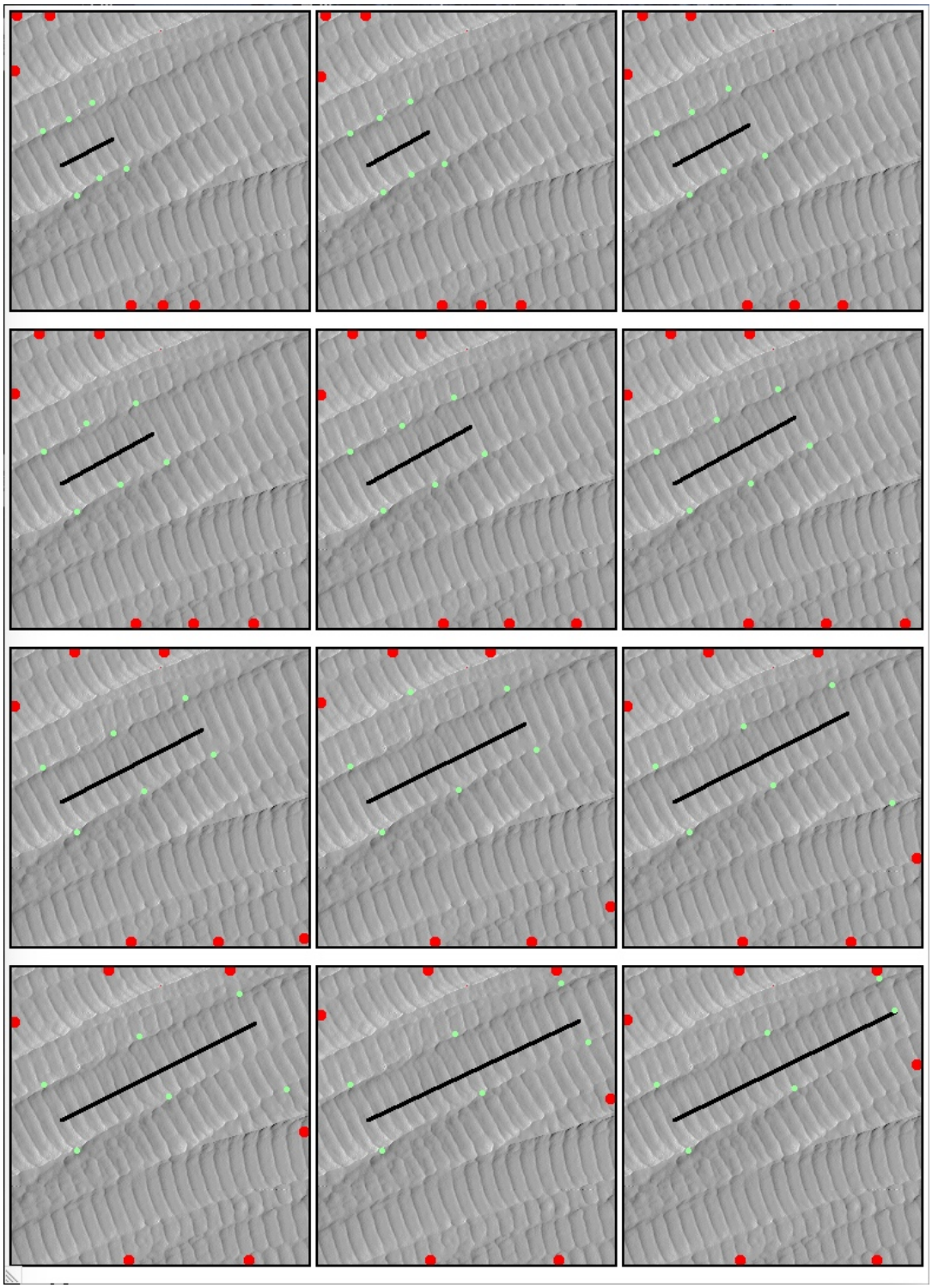


Figure 3.34: A ROI gradually elongated along the same fibre (fibre 1). (Read left to right and top to bottom.) The ROI origin is kept fixed. As the ROI centre and end positions move parallel to edge features as they appear and disappear, the algorithm succeeds and fails in finding the correct edge. (See green dots.) In this case, the fibre edges show sufficient contrast and most attempts by the macro succeed. 68 edge points were identified correctly, 3 were identified incorrectly, and 1 was not found.

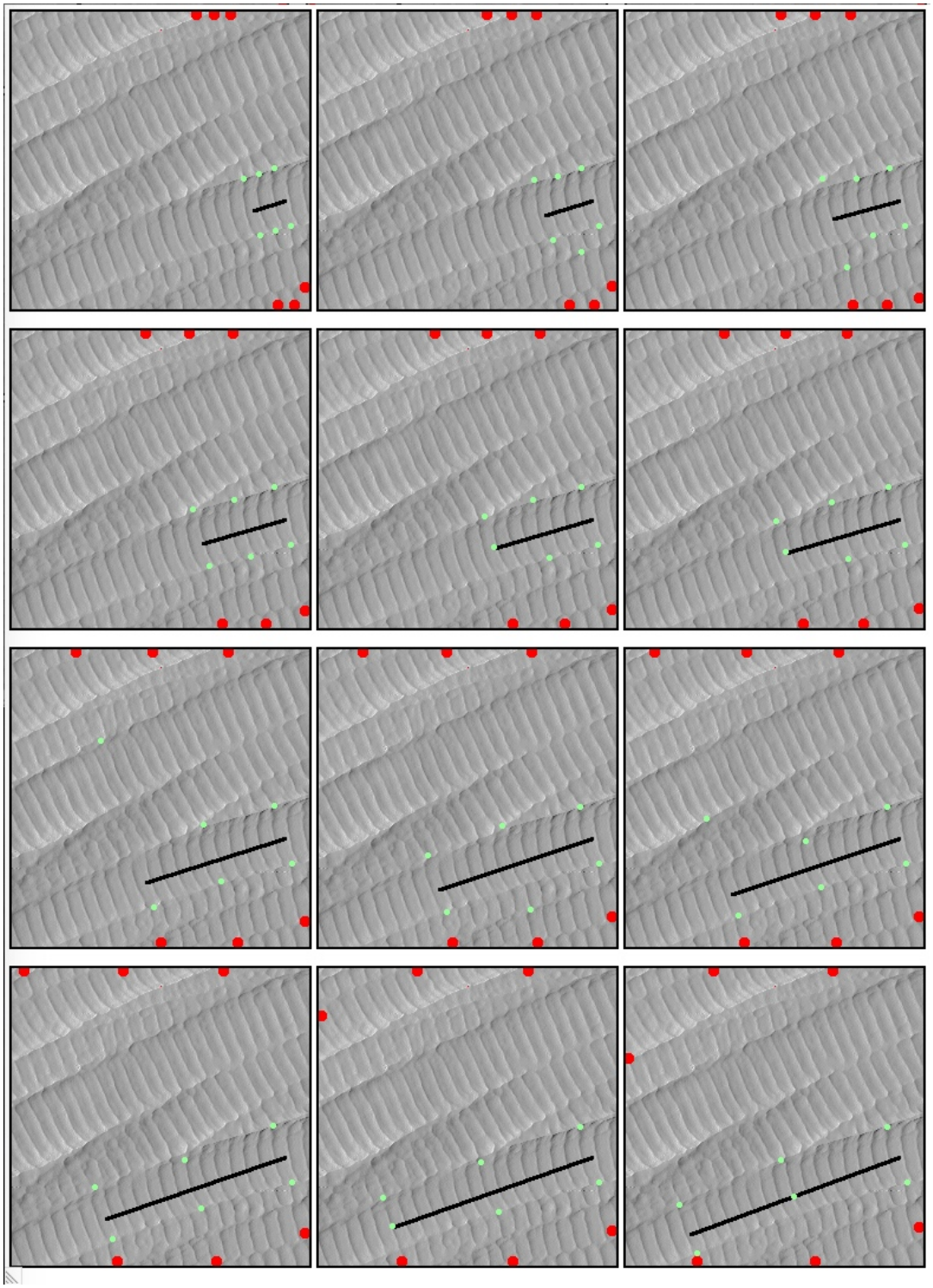


Figure 3.35: A ROI gradually elongated along the same fibre (fibre 2). (Read left to right and top to bottom.) The ROI origin is kept fixed. As the ROI centre and end positions move parallel to edge features as they appear and disappear, the algorithm succeeds and fails in finding the correct edge. (See green dots.) In this second case, the fibre edges often are too faint and cannot be identified by the macro. 62 edge points were identified correctly, 6 were identified incorrectly, and 4 were not found.

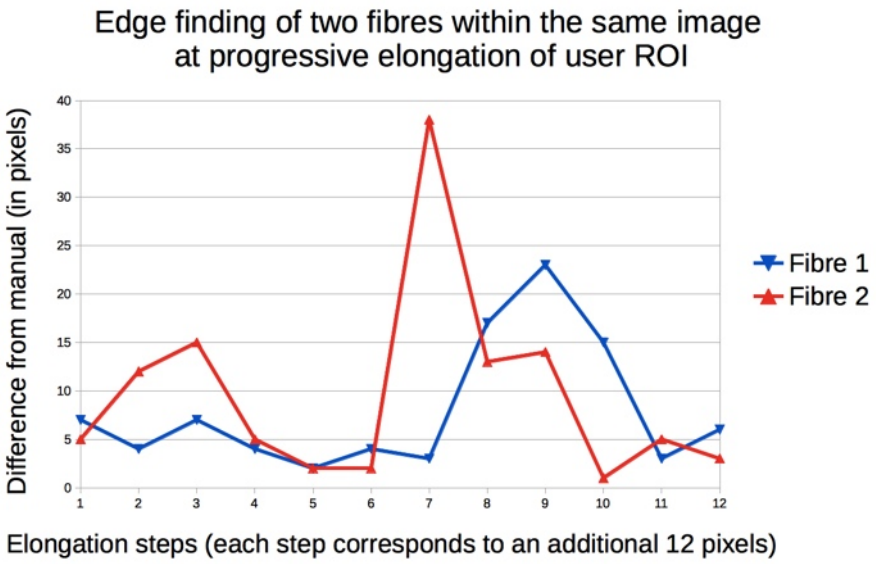


Figure 3.36: Difference between the average diameter measured by the macro and the manually measured diameter, plotted for each elongation step. At each step, the environment of the points at which the secondary ROIs are drawn alters and leads to varying success in edge finding. The average deviation from manual measurement was 7.9 pixels in the case of fibre 1, and 9.6 pixe;s in the case of fibre 2.

but once the variability of the user ROI is addressed, it will be possible to measure a large number of fibres with minimal effort.

3.5 Future work

As demonstrated above, the most important aspect of future work will be to remove the algorithm's sensitivity to the exact position of the first ROI. This can only sensibly be done when the algorithm is implemented in a compiled programming language, like PASCAL in the case of IMAGE SXM.

Several improvements to the *findstep* procedure could be envisaged. At present only a small portion of the fibre edge along the user ROI is sampled, but it is feasible to extend this portion by drawing multiple secondary ROIs at several fixed locations. Alternatively, secondary ROIs could move pixel by pixel along the initial ROI, thus substantially increasing the measured edge lengths. With a larger number of values for edge points it would be much easier to eliminate outliers and arrive at a more reliable final measurement, especially as there is a tendency of the macro results to be either reasonably similar to manual measurements or very different. With a larger number of measurements it would also be possible to apply statistical tests within every run of the macro and set custom criteria for whether a value is discarded or accepted. With an ability to autonomously assess the quality of the measurements, the macro could run through a range of parameters by itself to find the optimal set for the current image. Such steps would all make the outcome less dependent on the user's choice and the quality of the fibre, and therefore would be a substantial improvement to the current algorithm.

In addition to the angle-constrained Sobel filter, processing steps that further enhance contrast in the filtered image could be added. A wide range of image processing tools is available to prepare images for successful analysis, and any of these could be employed to improve the starting image, although their choice would most likely be very dependent on image type and analysis goals.

A general advantage of writing an algorithm in code, rather than macro language, and include it in Image SXM is that this would make the algorithm available to all users in future distributions of Image SXM. In this way it would also be possible to refine the output and combine separate text output files into a single results file.

In the present macro, the user has to find the relevant frequencies of the line ROI FFT by placing the cursor over the plot and reading the frequency in the info window, or by looking up the frequencies in the saved text output file. The reason for this is that the array holding the frequency data is not accessible from the macro language. Implementation in pascal would allow a procedure for finding the maximum frequency to be added as part of the algorithm.

Additional functionality could include the possibility for the user to draw more than one ROI at a time in one or more images, allowing better use of the operator's time. The current requirement for a line ROI could be replaced by a freehand, or polygonal ROI to deal with curved fibres, although especially in these cases care would have to be taken not to interpret real variations in fibre diameter as failure to find the right value.

While in this work only a small number of images of collagen fibres in the sclera obtained by AFM were studied, it is possible that the same or a similar algorithm could be used in the analysis of very different images. Images of other biological structures may show similar image characteristics, but also non-biological images may benefit. Analysis of the diameter of long, rope-like structures could be of interest in many fields, and as the real scale of the objects imaged is irrelevant, this could apply to fields as diverse as engineering, food science, geography, textiles or medicine.

3.6 Conclusion

In this chapter, we develop an algorithm requiring only minimal user interaction, which allows measurement of the diameter and banding frequency of collagen fibres in AFM images with high success rate. We demonstrate its robustness in the face of a wide range of values for all three analysis parameter and show its overall reliability. To our knowledge, such measurements of collagen fibres have so far only been carried out by visually reading off distances, which is feasible for small numbers of measurements. Our method, however, allows gathering a large number of measurements quickly and under identical conditions, with lower potential for user error, making a much higher throughput possible.

Due to its robustness, the same algorithm can be applied also to structurally related image data of other subjects (see 7).

Chapter 4

Polarised Light Microscopy

4.1 Principles of Polarised Light Microscopy

4.1.1 Polarised Light

Polarised light microscopy (PLM) is based on the principle that natural light can be converted into polarised light and that certain materials treat light differently depending on its polarisation state. A light beam can be thought of as being made up of a number of individual light waves, with each light wave oscillating within a specific plane. Light, as a kind of electromagnetic radiation, has an electric and a magnetic field vector, and it is these that describe the oscillation. When a single light wave propagates in a certain direction, the electric field vector (e-vector) oscillates at a certain frequency in a plane perpendicular to this direction, while the magnetic field vector oscillates perpendicularly to both the electric field vector and the propagation direction (fig. 4.1). As any vector, the e-vector can be seen as the sum of two vector components at right angles to each other. This is important for understanding polarisation.

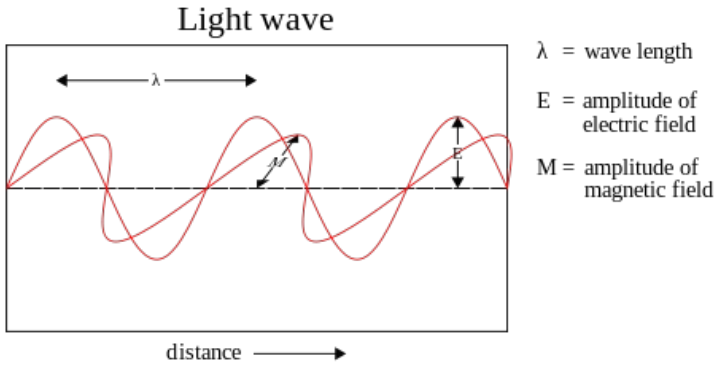


Figure 4.1: Light represented simplified as an electromagnetic wave, in which the electric field vector oscillates at a certain frequency and amplitude in one plane, and the magnetic field vector equivalently in a perpendicular plane. (Adapted from [101].)

When many such light waves with identical orientations of the electric and magnetic field vectors make up a light ray, this light is said to be plane or linearly polarised. For practical reasons, only the e-vector is usually taken into account, and the direction of the e-vector is called the axis of polarisation.

Natural sunlight contains many such single light waves with no preferential axes of

the e-vectors, and therefore is regarded as unpolarised (or randomly polarised). Usually, plane polarised light is generated from unpolarised light by filtering out and eliminating all but one polarisation direction. This can be achieved artificially in a number of ways, such as sending the light through an optically anisotropic material (a material whose optical properties vary with direction), or through reflection on surfaces. As this filtering is a scattering phenomenon, the interaction of the light with electrons in the material is crucial.

Anisotropic materials that are used for generating linearly polarised light (and in this function are called polarisers) need to be arranged on the nanoscale, either through their crystal structure or their molecular arrangement, in such a way that their electrons show appropriate spatial arrangements to treat light with different polarisation direction differently. Materials whose electrons are free to move in one direction, but not in others, interact with unpolarised light in such a way that the elements of the field vectors that are in line with the direction of free movement of the electrons are absorbed, whereas the elements perpendicular to this are not. Modern polarising filters consist of long stretched-out films of polymers that meet these criteria. Their electrons absorb the e-vector components parallel to the main axis of the long polymers, but not the perpendicular component. Thus light shining through such a material becomes linearly polarised with a polarisation plane at right angles to the orientation of the polymer strands. Before the advent of synthetic chemistry, only a few, rare minerals were suitable for this purpose, due to their crystal structures. Tourmaline is one such example.

Reflection on surfaces is a prominent source of polarisation in nature. A well-known example is the glare of reflected light on bodies of water or glass, which can be removed from photographs by the use of polarising filters. When an unpolarised light ray hits a smooth surface, the e-vector components parallel to the surface get reflected, whereas the components contained within the plane of incidence either get refracted or absorbed. Thus, the reflected (and also the refracted) light beam become partially polarised. The degree to which this happens depends on the refractive indices of the materials on both sides of the surface and the angle of incidence. The angle at which all of the light becomes polarised is the Brewster angle. The value of this angle depends on the materials, since its tangent is the ratio of the refractive index of the surface material over that of the material from which the light enters.

Materials not only reflect and absorb light differently depending on the relative directions of the e-vector, but can treat them differently in other ways, too. There are substances which intrinsically have different refractive indices along different axes, and are thus called birefringent. The directions along which the refractive index is smaller and larger are called the fast and slow axis, respectively. When light enters such a birefringent material in the same plane as these two axes, the beam is split up. When the component of the e-vector points towards the slow axis it gets refracted more than when it points towards the fast axis. This causes the two parts to take a different path through the material. Some crystals have long been known for this property. Calcite is a well known example, which appears to duplicate text when viewed through a sufficiently thick crystal.

When linearly polarised light enters a birefringent material at an angle perpendicular to the plane defined by the fast and slow axes, the components of the e-vector parallel to the slow axis experiences a different refractive index from the e-vector component parallel to the fast axis. This leads to a phase shift between the slow and fast axis components, which means that the slow axis component gradually starts lagging behind the fast axis component. The curve that the e-vector describes then changes from a line (linearly polarised) to an ellipse (fig. 4.2) and the light is called elliptically polarised. A special case exists when the thickness of the birefringent material is such that the phase difference between the two components is exactly $\pi/2$. At this point, the e-vector describes a circle and the light is called circularly polarised (fig. 4.3). If one could probe the polarisation state of a light beam throughout the thickness of a uniform, birefringent material, the state would be seen to oscillate through linearly (phase shift $= 2\pi n$), elliptically (phase shift $\neq 2\pi n$ and $\neq n\pi/2$) and circularly polarised (phase shift $= n\pi/2$). (Linear polarisation too can be regarded as elliptical, with the special case of a flat ellipse.) As refraction is wavelength-dependent, the thickness at which these conditions (linearly, elliptically and circularly polarised) are observed are different for each wavelength. Thus, when constructing a filter for converting the polarisation state of light (a so-called wave plate), a suitable thickness of the birefringent material would have to be chosen for each particular wavelength of interest. In practice, however, green at 500 nm is usually used as a compromise, since it is in the middle of the visible spectrum. This is the principle behind most such filters, which confer a specific phase shift to the e-vector components. In half wave plates this shift is π, in quarter wave plates it is $\pi/2$. As the difference in phase shift between the chosen and other wavelengths increases with material thickness, it is on the one hand desirable to keep the thickness of the wave plate minimal, within the constraints of the $= 2\pi n$ wavelength steps. On the other hand, however, in practice higher order wave plates are used, since they are more robust and can be handled more easily.

4.1.2 The Polarised Light Microscope

Polarised Light has been used for microscopy purposes at least since the early twentieth century, and is thus a well established light microscopical techniques ([103]). A simple polarised light microscopy setup (fig. 4.4) consists of a conventional light microscope with a linear polariser and an analyser inserted in the light path. (An analyser is another linear polariser, but on the opposite side of the specimen). Unpolarised light enters the polariser and only light with one polarisation direction is transmitted. This light enters the first quarter wave plate, which is positioned so that the fast and slow axes are both at 45 degrees to the polarisation direction of the polariser. The e-vector has an equal component along the fast and slow axis of the quarter wave plate and the light becomes circularly polarised. Then the light path crosses the specimen, followed by the second quarter wave plate positioned at 90 degrees to the first one. This converts the circularly polarised light back into linearly polarised. The last step is the analyser, oriented at 90 degree to the polariser, and at 45 degree to the quarter wave plate. If no optically active specimen is inserted in the light path, the analyser extinguishes all light in the path, since

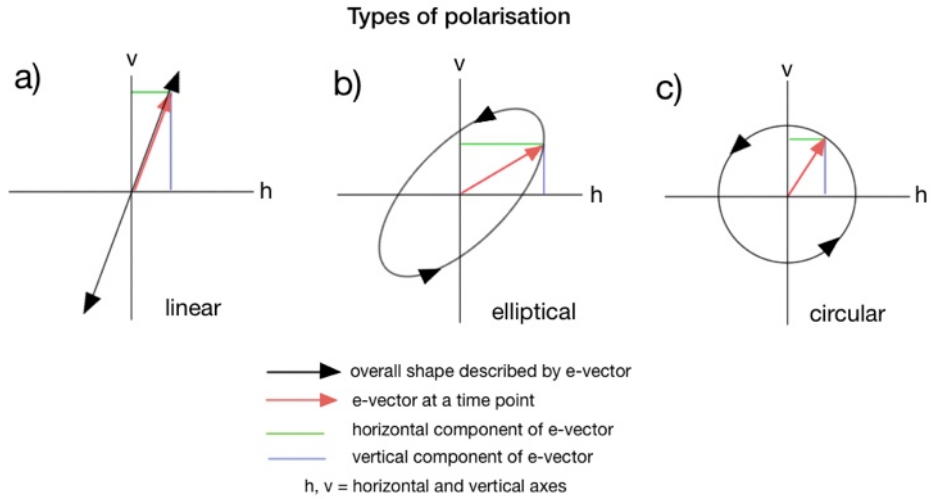


Figure 4.2: The principal polarisation types. The light is assumed to emerge at right angle from the page in the direction of the viewer. In the case of linear polarisation (a), the e-vector telescopes back and forth in one direction. The ratio of the compound vectors remains constant. In the case of elliptical polarisation (b), the e-vector describes an ellipse. The ratios of the compound vectors change with the phase. In the case of circular polarisation (c), the e-vector describes a circle. The ratios of the compound vectors change with the phase, but the vector sum always adds up to a constant value.

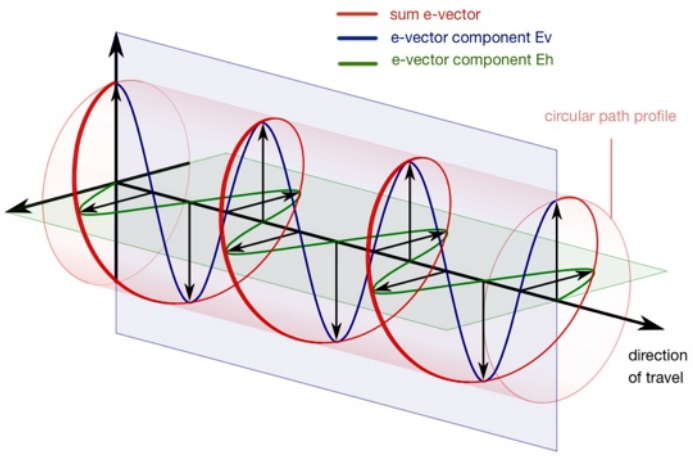


Figure 4.3: Circularly polarised light. In red: Path described by the sum e-vector; in blue: path described by the vertical component of the e-vector; in green: path described by the horizontal component of the e-vector. This is a special case of circularly polarised light, in which the phase shift between the two component vectors is $\pi/2$. (Adapted from [102].)

it is at right angle to the first polariser. When a birefringent sample is placed between the two quarter wave plates, however, the e-vector components get altered by the specimen, so that the light is no longer perfectly circularly polarised and a part of it can escape the analyser.

Placing two quarter wave plates on each side of the sample ensures that light is detected no matter which angle the sample is oriented. Without this measure, sample areas that absorb the e-vector in the plane of the polariser would appear dark, and areas that produce linear polarised light with an e-vector perpendicular to the analyser would be extinguished. An example of this is the Maltese cross pattern typical of starch grains in a setup without quarter wave plates.

This is the simplest setup, but other setups using different combinations of wave plates, as well as more advanced modifications have been developed. Their description is beyond the scope of this thesis.

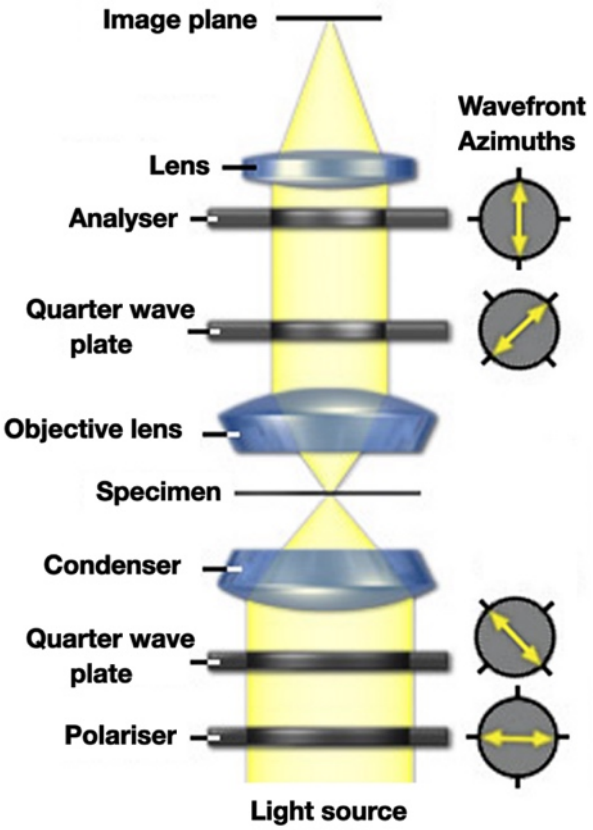


Figure 4.4: Schematic drawing of the light path through a simple polarised light microscope. (Adapted from [104].)

Historically, PLM has been widely used in geology, to examine thin sections of rocks, and in biology, to examine various birefringent structures of living matter. While its use in geology has mostly survived, it has largely fallen out of fashion in biology. Only a very small number of articles on straight polarised light microscopy have been published in the last few years in the life sciences. (Polarisation in the context of nonlinear microscopies, however, has seen more publications.) New microscopy methods are assumed to have superseded conventional light microscopy techniques, and many experts in PLM have long

retired without training successors. Contrary to the assumption that PLM is simplistic and outdated, polarisation is a complex phenomenon, and its use in microscopy is both potentially powerful and not trivial to interpret. Compared to the range of fashionable microscopies such as confocal, second harmonic generation, electron and scanning probe microscopies, it is cheap enough to be accessible to most scientists, and deserves far more notice than it is presently afforded. There are, however, still a small number of scientists who are engaged in developing this technique further and adapt it to special uses ([105], [106]).

4.1.3 Polarised Light Microscopy of Biological Materials

Many biological materials exhibit optical properties. Structural colours (colours originating from structural arrangement of tissue components, rather than from light absorption by pigments) are the cause of the spectacular iridescence seen in some animals. Certain animals can see linearly, and a few even circularly, polarised light, and some level of birefringence is almost ubiquitous in tissues. (A good review of optics in aquatic life is given in [8], on structural colours in general see [107], [108], and for circularly polarised light in nature see [109].) These kinds of optical activity are either due to the molecular properties of the individual biological molecules (intrinsic birefringence), or due to their supermolecular arrangements (form birefringence). The most abundant, fibrillar, collagens are found in long, axially aligned bundles and show marked intrinsic and form birefringence. In the human eye, and in vertebrate bodies in general, fibrillar collagens are the predominant source of birefringence. This property has been widely used in the past to microscopically examine collagen-rich tissues, mainly for histological purposes, though currently the use of polarised visual light for microscopic studies has largely fallen out of fashion in life sciences. In this study an additional staining method, picrosirius red staining, has been used that further enhances collagen's intrinsic birefringence by around 700% ([110]). This is achieved by largely electrostatic binding of the stain molecules to the collagen molecules, thus enhancing its intrinsic structural properties even further [111].

4.2 PLM Images and Analysis

PLM was used to acquire images from porcine sclera for the purpose of developing an image analysis strategy for collagenous and similar structures because of the easy availability of a polarised light microscope, porcine eyes and the simplicity of the staining method. The low magnification obtained from this light microscopic technique complements the higher magnifications achievable by AFM and TEM and allows a larger overview over the tissue structure. The thickness of the porcine sclera varied from around 0.5 to 1.5mm and the field of view in transverse sections usually covered the entire width of the thin areas, but only a fraction of the width in the thicker regions.

4.2.1 Preparation Methods

Pig eyes were collected from a local butcher, and a sagittal[1] slice from the middle of the eye was dissected. This tissue was then transferred into a 4% paraformaldehyde solution for fixation and the next day sent away to the Veterinary Pathology Laboratory at the University of Liverpool to be processed, sectioned and stained with a hematoxylin and eosin (H&E) stain and a picrosirius red stain according to a standard protocol ([112]). Section thickness was nominally 4μm, although this was not experimentally verified. The sections were then observed under circularly polarised light in an Olympus BX51 Transmitted and Reflected Light Microscope, equipped with two linear polarisers and two quarter wave plates, at 10 times magnification. Images were captured using a Colorview 3 camera and analySIS software. The red channel, which contained almost all the intensity, was then converted into an 8 bit grey value image and an analysis method was developed using the IMAGE SXM image analysis package ([50]).

4.2.2 PLM Images

Three porcine eyes were used for the development of this analysis, and six regions in ten locations (some regions were sampled on both sides) were imaged in each. As the structure and biomechanical properties of the sclera vary, the images were taken at anatomically different regions of the eye ball and at a magnification low enough to encompass a reasonable fraction of the width of the sclera. The dissection process is difficult in a tissue that is simultaneously tough and slippery, and thus maintaining consistency in the dissection is challenging. Also, positioning the tissue at exactly the right angle for embedding and sectioning cannot always be guaranteed. The fields where the images were taken are therefore approximate regions, rather than exact locations, and while every effort was made to obtain sections in a meridional plane, deviations from this were unavoidable, which led to an occasional increase in apparent scleral width. The regions examined were the cornea (C), limbus (L), medial region of the eye (M), posterior eye (P), area lateral to the lamina cribrosa (LC), and immediately at the optic nerve head (ON), as shown in fig. 4.5.

The field of view usually covered the entire width of the thin areas, but only a fraction of the width in the thicker regions. An example of an image from area M is shown in fig. 4.6. These original images were 5 megapixels in size and would have taken a prohibitive length of time to analyse by macro. For macro analysis image sections around one sixth that size were taken from a representative, central area of the original images.

The images show the characteristic red birefringence typical of picrosirius red staining of collagen ([110]). Whether it can be assumed that all birefringence in picrosirius red-stained tissues is derived from collagen has occasionally been debated. The answer depends to a certain extent on the tissue and the presence of other structures, such as mucus glands, which would also appear red under picrosirius red staining, but are not present in the sclera ([110]). It is accepted that a thick collagenous extracellular matrix makes up the main bulk of the sclera, and even if some birefringence did originate from other molecules, their contribution in this tissue would most probably be minimal.

[1] A sagittal plane is a plane that slices the body into right and left halves, and all planes parallel to this.

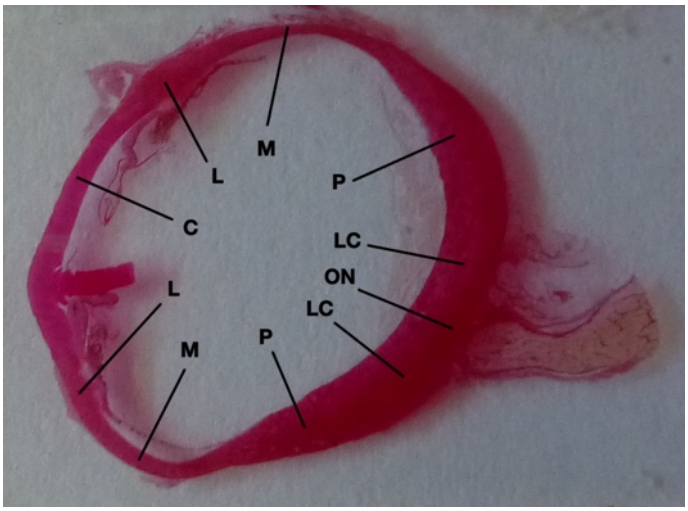


Figure 4.5: Example of section through a porcine eye ball stained with a standard histological stain (H&E). The areas where images were taken were the cornea (C), limbus (L), medial of the eye (M), posterior eye (P), lateral to the lamina cribrosa (LC) and optic nerve head (ON).

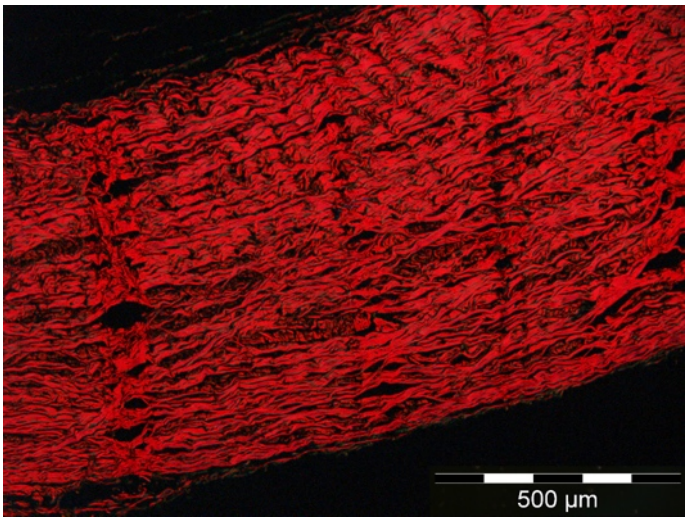


Figure 4.6: Example of PLM image at 10x lens magnification. Red staining is from collagen birefringence enhanced by picrosirius red. The scale bar is shown here for information and is the same for all images examined, but will not be shown in images used for analysis.

Image SXM converts colour images into three 8-bit images, for the red, green and blue components. When examined, there was little intensity in the blue and only slightly higher intensity in the green image. Neither of them appeared to show structure not also present at higher intensity in the red channel, and therefore only the red channel images were used. In total, 29 images were used for macro analysis.

4.3 Analysis Strategy

In geological images it is customary to glean information about the optic axes of specimens with the assumption of constant thickness and uniform crystallinity throughout the optical path. In the case of collagen, however, its optical properties appear to be affected not only by section thickness, but also by age, hydration state ([110]) and possibly also degree of packing and cross-linking ([98], [113]), all of which may vary throughout the specimen and could conceivably also lead to variations in local section thickness. It was therefore decided not to attempt to interpret any potential colour variation of the images in a quantitative way, but to use the birefringence as merely a tool for visualisation of collagen in the image. This approach also required a simpler microscopic setup.

As was the case with the AFM images, distinct bundles of collagen fibres in the PLM images appear present only until one actually attempts to follow a fibre bundle through the image. Multiple fragmentations and joins, as well as what seems to be fibres emerging at all angles, prevent every attempt to extract properties such as individual fibre length, thicknesses or angles. A more promising approach, therefore, than the pursuit of single fibre analysis, is the attempt to identify whole-image, textural characteristics, which may show typical patterns in groups of images. These ought to be related to the underlying tissue structure, but do not need to be identical with a structure at every image point. Such sets of characteristics that can be easily and quickly obtained by automated whole-field analysis may prove useful for diagnostic purposes. Larger sets of images may be analysed and characteristic properties identified to arrive at a classification. Then, unknown images in the form of, for example, biopsy or research samples could be compared to these previously obtained standards and classified as similar to healthy tissue, or compared to tissues with certain conditions. Furthermore, biomechanical studies could be supplemented by such whole-field characteristics, and simple models could be devised in such a way that the textural characteristics of biological tissue are mimicked.

The Autocorrelation Function (ACF) describes the correlation of an image with itself as a function of displacement. It calculates the correlation of the grey value of every pixel with the grey values of every neighbouring pixel as a function of the location of these pixels with respect to the first. The ACF can be visualised by imagining an image with a white circle on a black background and making two copies of this on transparent sheets (fig. 4.7). Each one is completely opaque, with the exception of a central, transparent circle. These two sheets are placed on top of each other on a light box. When the sheets match up perfectly, the centre circles overlap completely and the light transmitted through the sheets is brightest. As the sheets slide in relation to each other, the area of overlap shrinks, and less light can get through the remaining transparent area. The light passing

through the films at a given displacement and direction is then equivalent to the grey value of the ACF of the image at that displacement. This explains why the basic, maximum outline of an ACF of a simple, isotropic, black on white circle, is twice the diameter of the original circle. Properties like this allow conclusions to be drawn from the ACF about the original image, although in real, grey value images the situation in rarely as straightforward. In IMAGE SXM the ACF is calculated via Fourier Transforms, which restricts the image sizes to powers of two, but also allows fast calculation. While it is easy and straightforward to execute an ACF from the macro language, interpretation of the ACF is not easy. Heilbronner and Barrett [114] provide an excellent discussion of the use of the ACF in the context of Earth Sciences.

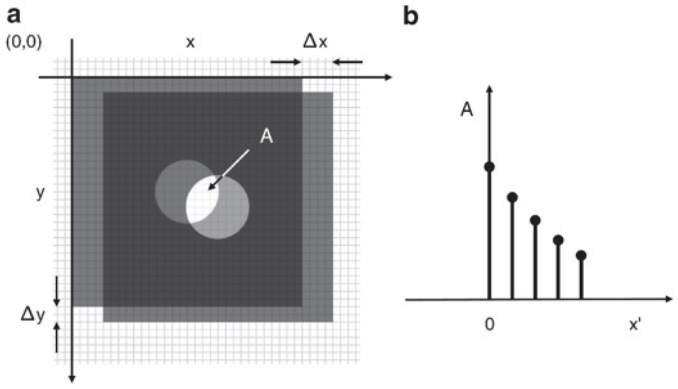


Figure 4.7: Visualisation of the ACF: (a) Two copies of an image are superposed, one copy is shifted in the x- and y-direction by Δx and Δy; A = area of overlap of circles; (b) the correlation between the image and its copy is measured as $A(x')$, the area of overlap as a function of x- displacement. Taken from [114], with permission.

The dimensions and units of the ACF image are identical to the source image, but pixel position in the ACF has the meaning of displacement, rather than position in the original image. The centre of the ACF is its zero point, and at this zero displacement the highest grey value is always found (in keeping with the analogy of light shining through two sheets when there is no displacement). Every position $(x\prime, y\prime)$ in the ACF is the end point of a neighbourhood vector from all pixels of the image. The grey value at point $(x\prime, y\prime)$ in the ACF represents the correlation of all original image points with all neighbours at the position $(x + x\prime, y + y\prime)$. Therefore, the ACF of an image shape is identical, no matter where in the source image the original shape is located. For simple shapes, the shape of the ACF replicates the original relatively closely, but for more complex shapes this is no longer the case (fig. 4.8). The size of the central ACF peak in any direction is twice the length of the shape in the original image in that direction, and the main orientation of the ACF follows the main orientation in the source image. When there are multiple shapes in the source image, though, this constellation of shapes can be regarded as a new shape in itself, and the ACF becomes more complex. In such a case not only the dark shape, but also the anticorrelated light areas become of interest, as they contain information about the spacing. Whether the grey values of the ACF around the central maximum go to zero at some point depends on the distances between the original shapes. If the original shapes

can be "slid against each other" in such a way that the overlap region disappears on one end before a new overlap develops on the other side (i.e. if the spacing is more than two diameters), the ACF's grey values do go to zero. If the spacing is shorter than that, the ACF does not have a zero baseline around the centre. The Particle Analysis function inbuilt in IMAGE SXM is convenient for extracting characteristics like area, angle, long and short axes, etc. from particles. Since a particle is by definition a contiguous black area in a binary image, for the Particle Analysis to be used the grey value ACF image has to be converted into a black and white image by thresholding. Where the threshold is set influences greatly the boundary of the central particle in the binary image (see below).

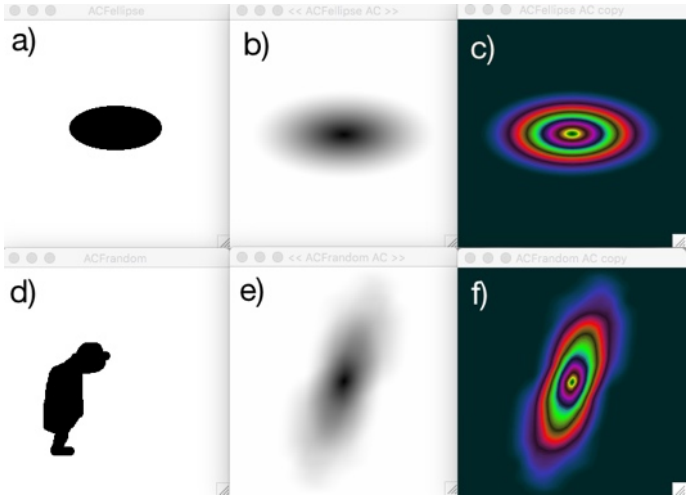


Figure 4.8: Simple shapes and their ACFs in grey scale and colour look-up table (LUT). A black on white ellipse as starting image in a) has an ACF whose contour lines are also ellipses, see b). This is more obvious when grey values are rendered in colour (c). In contrast, an irregular black on white image (d) has an only slightly anisotropic ACF, but does not replicate the starting shape contours (e), again shown more clearly in colour in f).

These properties of the ACF can be demonstrated for simple shapes ([114]), but real life images are often too complex for the ACF to be interpreted in such a relatively straightforward way. Fig. 4.9 shows an example of the difference between a real and an artificial image of long strands. In these images it is difficult to say which feature of the source image has caused which feature in the ACF.

In spite of such restrictions, what can be accepted is that the overall direction of the shapes in the source image will be reflected in the overall direction in the ACF, no matter how complex the source. It is also one of the great advantages of Fourier-based methods that complexity is no obstacle to analysis. As a quick (not shown) demonstration of the directional equivalence, when a real image is rotated through 90, 180 and 270 degrees, and the ACFs are thresholded, the Particle Analysis returns the same angle as in the original ACF plus 90, 180 and 270 degrees, respectively. Furthermore, long, thin and parallel fibres in the image will produce a long and thin ACF particle, whereas randomly oriented fibres or rounder shapes in the image will produce a more irregular shaped or rounded ACF.

As far as the detailed shape of the thresholded ACF in complex images is concerned,

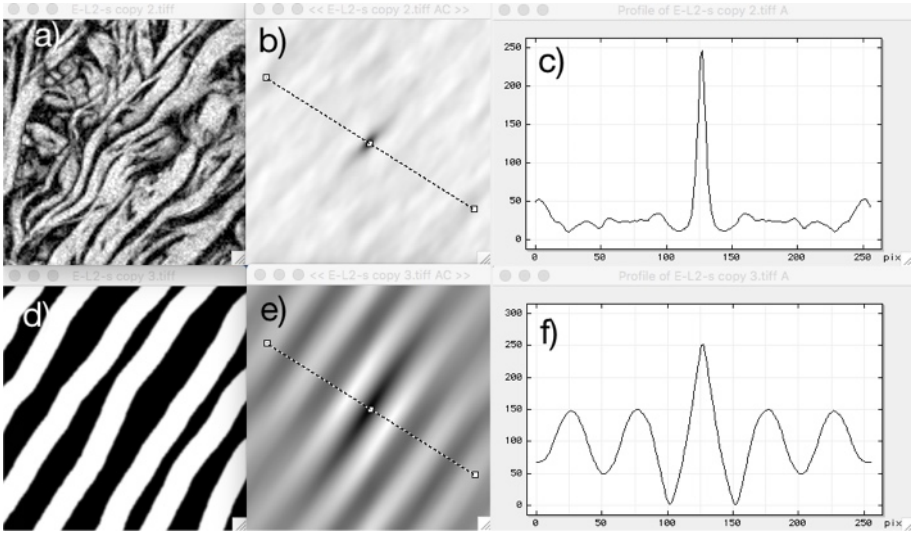


Figure 4.9: Example of the difference between the ACF of a real and a simplified image of similar shapes. a) real image; b) real image ACF; c) plot profile along line in b; d) simplified image; e) simplified image ACF; and f) plot profile along line in e. Note the difference in low contrast texture between the ACF images.

it too reflects but does not duplicate the overall shape present in the source image (i.e. the entirety of the image content, not the individual fibres). This behaviour is illustrated in figs. 4.10 and 4.11.

While in complex real images no direct inference can be made from the thresholded ACF to the exact individual shapes and directions that gave rise to it, the ACF can be used to characterise the whole image field, and provides a numerical way of comparing it to others. One of the advantages of the ACF is that no prior processing or segmentation of the image is necessary. This removes the danger of involuntarily interfering with image characteristics that would have been useful for characterisation, and also helps to make the technique more universally applicable.

PLM images of scleral collagen possess structure at many different length scales. As collagen fibre thickness and extent of branching vary in different areas of the sclera, and also lead to different biomechanical behaviour (see Chapter 2), characterising this tissue at different length scales could yield important insights for research, diagnosis and modelling. The principle of our analysis is therefore repeated at different length scales within the images.

4.4 Macro Development

The characterisation of the PLM images is based on the ACF and its Particle Analysis. Increasingly large square ROIs are moved over the image, row by row, in such a way that adjacent ROIs overlap by half their width and length. For each of these ROIs, an ACF is produced and then thresholded at a certain grey value. The resultant particle is copied into a blank image the same size as the starting image. This results in a new compound image (one per ROI size) containing an array of all the ACF particles, giving a visual

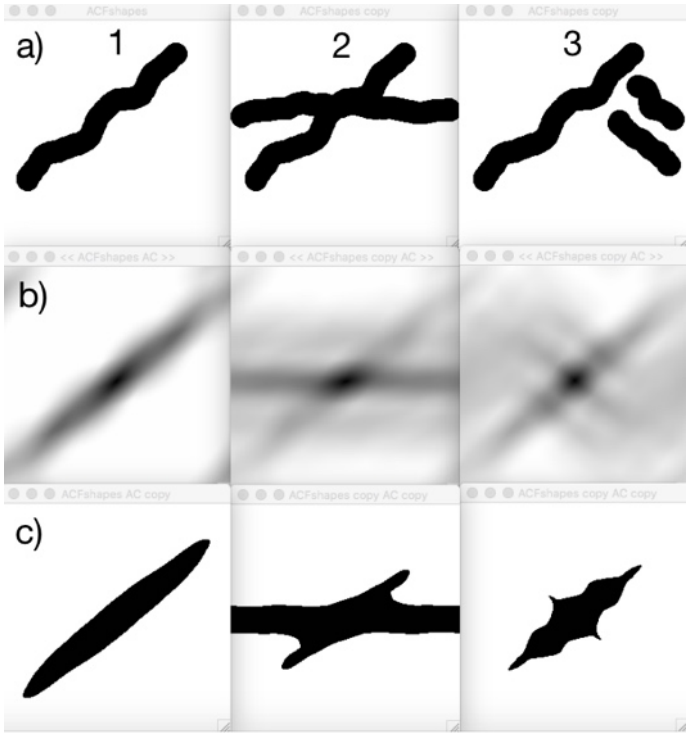


Figure 4.10: Example of how the presence of different shapes affects the shape of the ACF and its thresholded particle. Column 1,2 and 3 show different constellations of simplified shapes. Row a: the black and white shapes; row b: the ACFs of row a; row c: ACFs from b thresholded at a grey value of 83. The shape of the particle in row c, and how many are visible, depends on the chosen threshold value. See fig. 4.11.

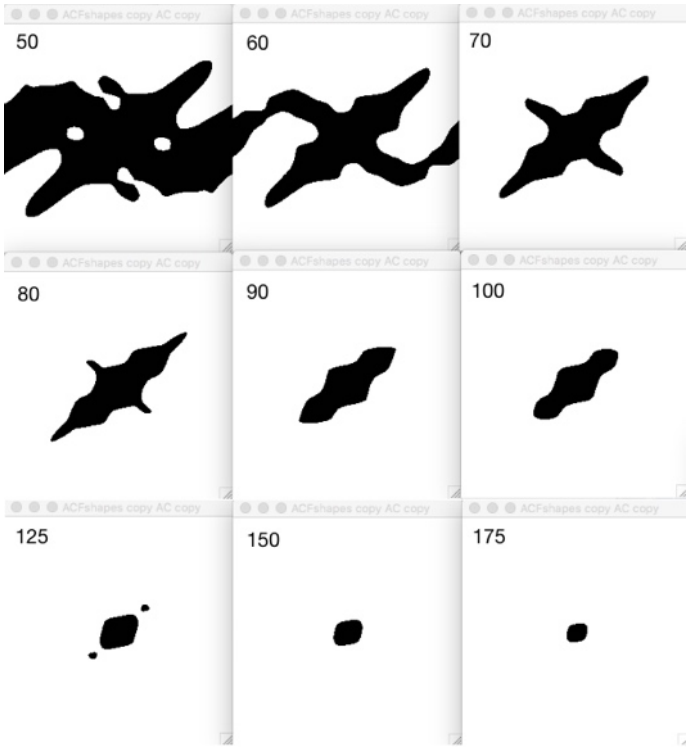


Figure 4.11: Dependence of the shape of the central particle of an ACF on the chosen threshold. The ACF 3b from fig. 4.10 was thresholded at the grey values given in the top left corners. There is strong dependence of particle shape on the threshold value.

overview over directionality in the image at that length scale (fig. 4.13). The Particle Analysis calculates area, angle and major and minor axes for each ROI, and the macro calculates the Axial Anisotropy (major axis divided by minor axis), as illustrated in fig. 4.12. These values are exported in a text file.

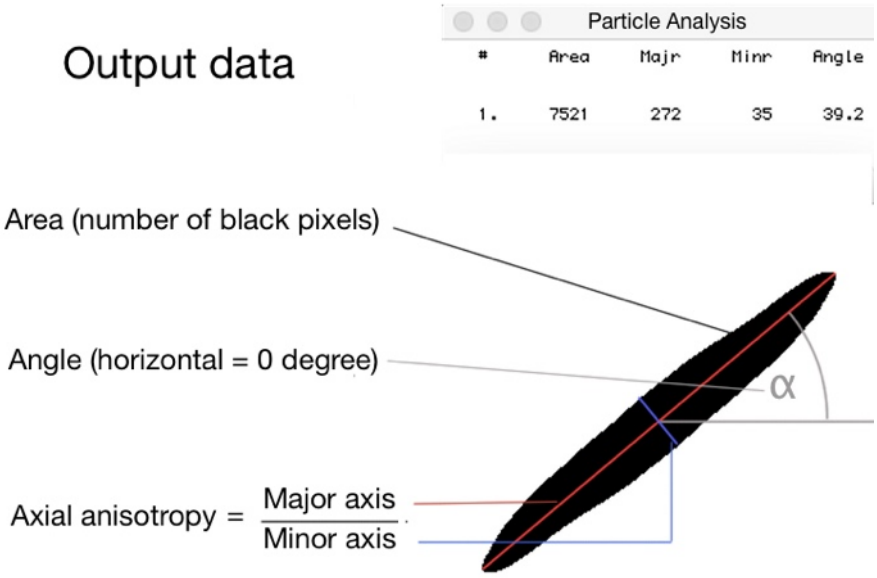


Figure 4.12: Output data from the Particle Analysis: area (black) is the sum of all black pixels in the central particle, angle (grey) measured counterclockwise and relative to the horizontal, major and minor axis (red and blue, respectively) are the longest diameter and its perpendicular drawn through the centre.

Since this is done for every ROI size (all ROI sizes have to be powers of 2, since the calculation is based on Fourier Transforms), changes in directionality for different ROI sizes can reflect variations in fibre direction or degree of intermeshing at different fibre sizes or length scales. Comparing ACFs in different areas within one image could also give information about regional variations.

Suitable length scales can be chosen by the user to account for different types of images and magnifications, since it is likely that the characteristic image features can be found at different relative sizes. However, due to finite pixel size, too small a ROI will encompass a section of the image that is too small to contain any useful information and introduce randomness in the ACF shape. This does not present a problem, though, as differences between images will simply not show at such scales. The largest sensible ROI size is the largest power of 2 that is smaller than the whole image size. Intrinsically, there will be a far higher number of output data at smaller ROI sizes than at larger scales, but as ROI sizes increase, each of these data points will be derived from a larger area.

4.4.1 Metrics

Of the four data types that the Particle Analysis returns in this analysis (area, angle and major and minor axis) only angle and major and minor axis are used directly for image

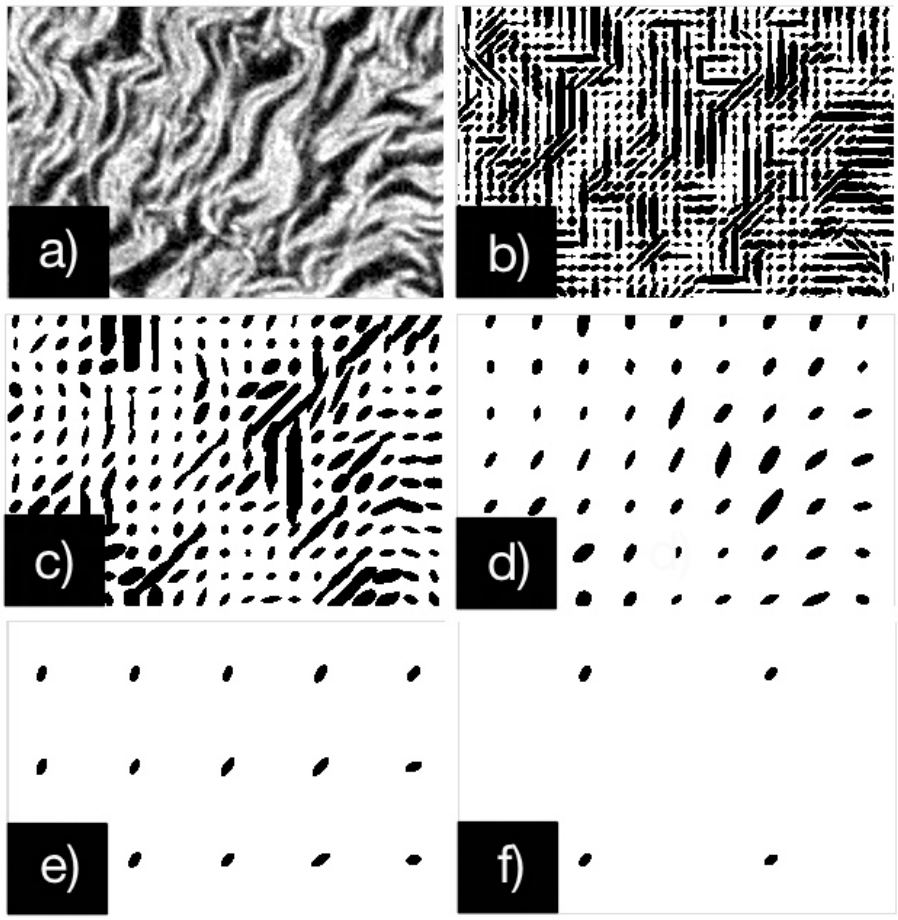


Figure 4.13: Example sections of the compound images of thresholded ACF particles. Section of original image (a) and ACF images of same section at ROI size 16 pixels (b), 32 pixels (c), 64 pixels (d), 128 pixels (e) and 256 pixels (f).

characterisation, whereas the area is used as a criterion for elimination of non-usable data points. When after thresholding an ACF a roundish particle with only a very small number of pixels is left, its angle and axial anisotropy are no longer meaningful. Thus, particles that combine an angle of zero with an axial anisotropy of 1 (round) and an area less than 10 pixels are rejected.

The Angle of the ACF particle can be used as a measure to estimate predominant direction of image structures. As discussed, this is the compounded direction of all structures present within the ROI. The values for the angle measurement from the Particle Analysis range from 0 to 180 degree. As angles are not linear but circular quantities, a mean or standard deviation cannot be calculated in the same way as for linear quantities. For example, two particles with angle measurements of 1 and 179 degrees will be almost identical, but their arithmetic mean would be 90 degrees, which is obviously wrong in this context. In order to adapt to this situation, angle measurements from the Particle Analysis were first multiplied by 2 to scale them up to 360 degrees. These angles were then regarded as unit vectors (with length = 1) and added (fig. 4.14) over the image. The sum vector's angle was calculated via the arctan of its x and y components, $\delta = \arctan(y_s/x_s)$, and this angle was divided by 2 to return to the correct value. This gave the *Average Angle*. The length of the sum vector s was calculated via Pythagoras's theorem from $s = \sqrt{x_s^2 + y_s^2}$ and is a measure for the divergence of the angles, similar to the standard deviation for linear quantities. If all n added angles point in the same direction, the sum vector has the length of $n * unitlength$, which is its maximum possible length. The more the added unit vectors diverge, the shorter the length of the sum vector. When, in analogy to the standard deviation, the length of the sum vector is divided by n, its maximal length becomes 1. A random distribution of a sufficiently large number of angles would therefore yield a length of 0, and the more aligned the unit vectors are the more this value approaches 1. This standardised length of the sum vector was given the abbreviation *AngleR* and was used as a measure for variability in the angle data, ranging from values 0 to 1.

AngleR is the first metric used in the analysis.

Comparing the *Average Angles* between different ROI sizes in theory allows correlation of predominant direction with length scale. In the eye, and especially at low magnifications as those used in this analysis, however, the curvature of the eye ball introduces an intrinsic variation in tissue directionality within one field of view, even if all fibres lay in perfect concentric layers. The absolute angles in the field of view furthermore depend on the orientation of the section under the microscope, and in non-perfect sections it is often not possible to determine which way to position the microscope slide to achieve alignment. For these reasons, while the *Average Angle* may be a useful quantity in itself for characterisation in other tissues, in this analysis more attention was paid to the derived quantity of *AngleR*, which is likely to have more relevance due to its ability to give information on parallel or random fibre directions. It should be noted though that the *AngleR* has an intrinsic tendency to slightly increase in value with ROI size, since with small ROI sizes far more angles are added up, with a higher propensity for randomness than with larger, less numerous ROIs.

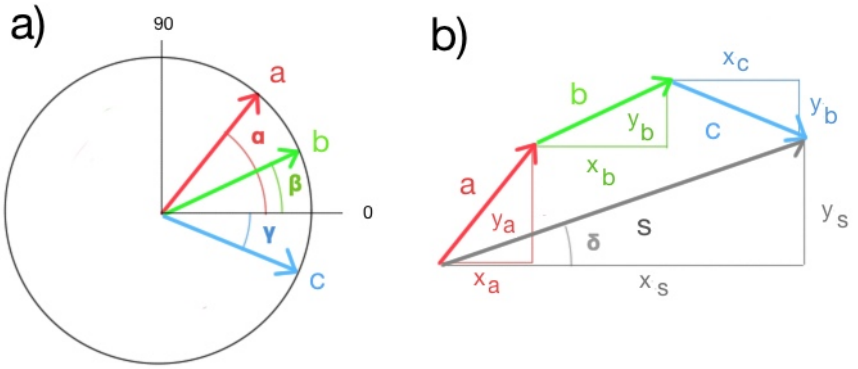


Figure 4.14: Calculation principle of the sum vector's length and angle. a) Angles α, β and γ are measured angles, already multiplied by 2. a,b and c are the corresponding vectors. b) Vector addition results in a sum vector with length s and angle δ. The *Average Angle* is calculated as $\delta/2$. *AngleR* is calculated as $s/3$, with 3 vectors added.

The axial anisotropy (a quantity we called *Axanis*) can also broadly be related to the intermeshing of collagen fibres. When all fibres run parallel to each other, the ACF will reflect this in a long and thin particle and this will yield a high value for the *Axanis*. When there is a high degree of intermeshing, fibres will run in many directions, and the image section is likely to contain smaller fibre segments. The ACF particle will be shorter and tend more towards roundness, and the *Axanis* will be closer to 1. It also has to be born in mind that fibre structure is not the only thing influencing the *Axanis*. It has been shown (fig 4.11) that the level at which the ACF is thresholded affects the particle shape, and thus the *Axanis* of the particle. The relative sizes of the image structures to the ROI size also influence the *Axanis*. Different ROI sizes, even if thresholded at the same grey value, will result in a differently shaped particle (since a different subarea is captured within ROIs of different sizes) and therefore a different *Axanis*. These factors should be kept in mind when interpreting the *Axanis* results. The values of the *Axanis* theoretically range from 1 to infinity, but in practice the maximum value is limited to $\sqrt{2}$ ROI size. The average of the *Axanis* values of all the ROIs of one size over the image is called *AxanisAve*.

AxanisAve is the second metric used in the analysis.

The standard deviation of the *Axanis* is called the *AxanisDev*. It describes how much variation there is in particle shape. When the *AxanisDev* is calculated from the particles of the whole image at one ROI size, it can indicate the variation in *Axanis* within the same or between different image regions at one length scale. If the *AxanisDev* is significantly different between ROI sizes, this is an indication that the variation in the *Axanis* (and the underlying image structure) changes at different length scales.

AxanisDev is the third metric used in the analysis.

All these metric quantities produce one value per ROI size per image.

Table 4.1 summarises these values.

Metrics	Definition	Related to
AngleR	Standardised length of the sum of the unit vectors derived from the angles of all ACF particles at each ROI size	How much the fibre directions diverge over the whole image or if there are regions with fibre directions parallel within the region but different between them, at each ROI size
AxanisAve	Axial anisotropies (major/minor) averaged over the ACF particles of the whole image at each ROI size	How fibrous the image texture is, or how parallel the fibres are on a length scale smaller than the ROI size, at each ROI size
AxanisDev	Standard Deviation of the axial anisotropies of the ACF particles of the whole image at each ROI size	How homogeneous the fibre texture is, or how parallel the fibres are on a length scale smaller than the ROI size, at each ROI size

Table 4.1: Brief overview of the metrics used in the analysis, *AngleR*, *AxanisAve* and *AxanisDev*.

4.4.2 Macro outline

The macro for PLM ACF analysis starts at the smallest user-defined ROI size and after completing this loop moves on to the next larger, and so on until the largest user-defined ROI size is completed. The loop consists firstly of the generation of a new, initially blank image into which the thresholded ACF particles will be copied. Then, row by row and column by column, a ROI is drawn into the image and copied into a new ROI-sized window, where the ACF is applied and thresholded at a constant grey value of 150 (determined by prior optimisation). The Particle Analysis is carried out and measures the angle of the central particle, the particle area in number of pixels and the length of the major and minor axes. The *Axanis* is calculated by the macro and these values are entered into an array. The ACF particle is copied and pasted into the same ACF window location as the ROI was in the original image. The ROI-sized window is then disposed of and in the main image the ROI moves on by half its length. The process is repeated until the remainder of the width and height of the image no longer allows the placement of another ROI. The data are written from their arrays into a data text file, which is saved and closed. The ACF Image is also saved and closed. This completes the analysis of one image, but if several images are being analysed, then the macro automatically moves on to the next image and repeats this process. A sketched outline of the macro procedure is given in fig. 4.15.

4.4.3 Procedure *ACFtilespart*

This procedure carries out the entire analysis proper. Its function is described in the outline of the macro as a whole (see section 4.4.2).

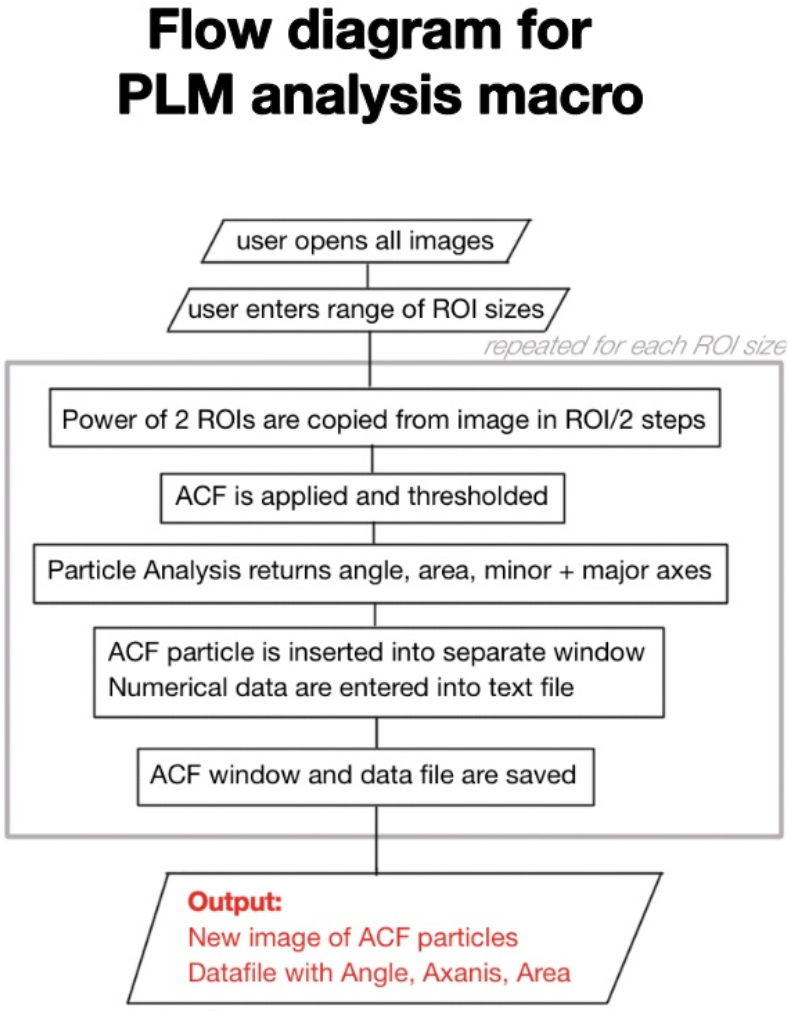


Figure 4.15: Brief outline of the macro for PLM analysis. The user input is minimal with decisions only about the range of ROI sizes. Although this macro has a small number of separate procedures, all the main parts listed here are carried out by procedure *ACFtilespart*. Auxilliary procedures are not mentioned here.

4.4.4 Procedure *ACFtileswROIsizes*

This procedure is auxiliary to the main function. *ACFtileswROIsizes* controls the cycle through increasing ROI sizes and calls the procedure *ACFtilespart*.

4.4.5 Macro *ACFtileswROIsizesallimages*

Although this is the macro proper, it too is merely auxilliary. It requests the information about the range of ROI sizes and number of images to be analysed from the user and then calls the procedure *ACFtileswROIsizes*.

4.5 Macro Evaluation

4.5.1 Impact of ROI size

The size of the ROI has an important impact on the size of the image features that can meaningfully be measured in it. A ROI that is small in relation to the relevant image feature can give no meaningful information of what kind of larger structure the image contains, or what texture these structures are part of in the image. On the other hand, if the ROI size is large in relation to the image features, information of their arrangement can get lost when there is regional variation on a smaller scale than the ROI. For example, a large ROI may contain fibre bundles showing internally strong, but regionally diverse, directionality. In this case the ACF may show little axial anisotropy, and any measured angle will be influenced by all the various angles of the fibre bundles in the ROI. In fig. 4.16 the ACF of the large ROI (b) shows much smaller axial anisotropy than the ACF of the small ROI (inset in b) from a central part of the larger ROI, and the particle angles are substantially different.

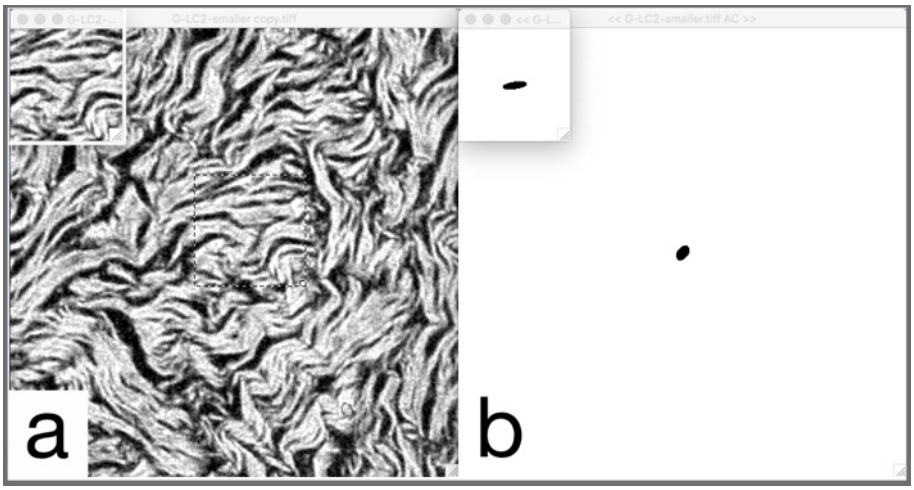


Figure 4.16: The large ROI (a) shows little axial anisotropy in its thresholded ACF (b), however a smaller ROI (inset in a) from the centre of the image shows a much more anisotropic central particle. Threshold in both cases is gv = 74.

For images that contain specific structures at distinct length scales, as the image gets segmented into ROIs of increasing size, the ROI sizes may move from being too small to

being a good match for the length scale of the structures to becoming too large to resolve details in the structures. In the case of images that contain fibrous structures a sign that the ROI size is sufficiently large might be the appearance of larger axial anisotropies. However, if on larger scales the fibrous texture became more randomly oriented, the *Axanis* would shrink again, whereas if small fibres on larger scales merged into larger, parallel bundles, the *Axanis* would continue at a higher value. To recognise and correctly interpret the meaning of the quantities measured for different kinds of complex images, the experimenter will likely require some familiarity with the images and experience with the applied method. This discussion relates mainly to the interpretation of the behaviour of the *AxanisAve*, *AxanisDev* and *AngleR* at different length scales in order to draw conclusions about single-image features. However, this is not necessary in order to use this analysis to classify different images or image groups. Identification of the *AxanisAve*, *AxanisDev* and *AngleR* with distinct image features is not necessary for finding common patterns in these metrics in larger data sets.

To see how the measurements in a simple dummy image may be affected by different ROI sizes, fig. 4.17 was analysed at ROI sizes 16 to 64. The lines were drawn at a 30 degree angle. At a small ROI size, the parts of the image feature captured by one ROI changes as the ROI moves across the image (fig. 4.18), resulting in periodically changing ACFs. This is particularly obvious where the distance between the lines in the image is varied slightly. As ROI size increases, the image parts captured by one ROI become more similar as a greater part of the image is sampled each time, and the differences even out. This sampling artifact can still be present but will be less noticeable in real images. The analysis of a real image would furthermore involve a larger data set, and the average values would likely not be influenced to a significant extent. Even in the small example shown here, the average for the Angle measurement is very close to the true value (fig. 4.19). Nevertheless, this is another reminder that caution must be taken when attempting to use only a few small-ROI ACFs for extracting image characteristics.

4.5.2 Testing Thresholds

In order to investigate the sensitivity of the results to the grey value at which the ACF is thresholded and to determine a suitable threshold, a range of threshold values were tested.

To ensure the results would apply not just to one image, two images were used, one with slightly denser fibrous structures (G-LC2) than the other (E-L2), see fig. 4.20. Three ROIs were arbitrarily selected by hand in each image for each of the ROI sizes from 16 to 1024 pixels. The ACF was applied and each particle thresholded at grey values between 60 and 240 (in steps of 10), followed by Particle Analysis. The resulting data were collated into three graphs, for *AxanisAve* (fig. 4.21), *AxanisDev* (fig. 4.22) and *AngleR* (fig. 4.23).

Each graph point consists of up to six data points per image. There was a cut-off between the threshold values from 100 to 200, which all yielded very similar results, and the lower and higher thresholds, from 60 to 90, and from 210 to 240, respectively, which showed somewhat erratic behaviour. This effect is due to the fact that when particles become very small at high thresholds, their measurements can change drastically by the

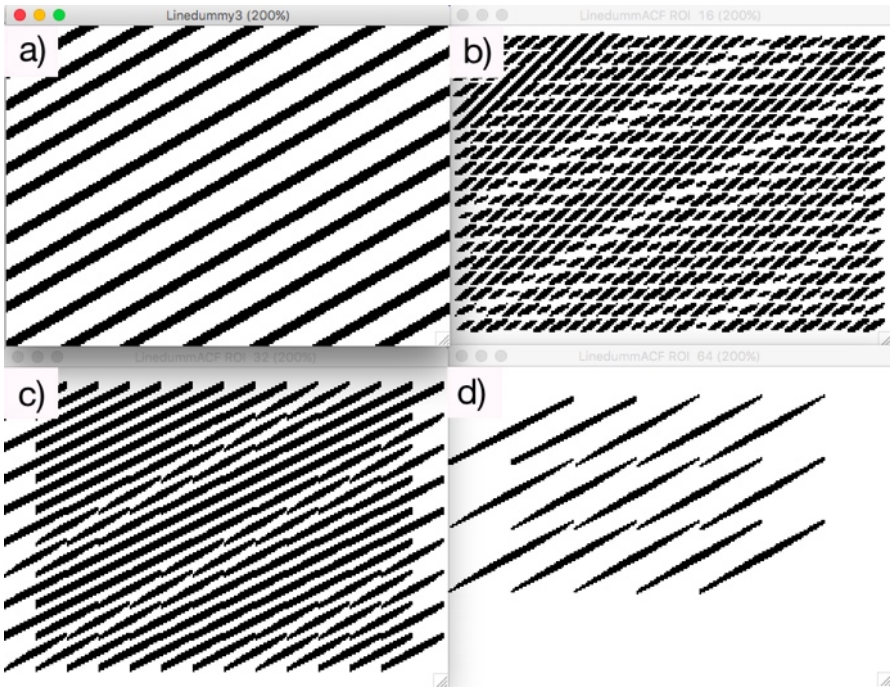


Figure 4.17: Example of sampling artifacts from a linear dummy image. A manually created dummy image with parallel, but not completely equidistant, lines at 30 degrees (a) and ACF particles at ROI size 16 (b), ROI size 32 (c) and ROI size 64 (d). Note the variation in the thickness, angles and areas, greatest at ROI 16, but still present to a lesser extent in ROI 32.

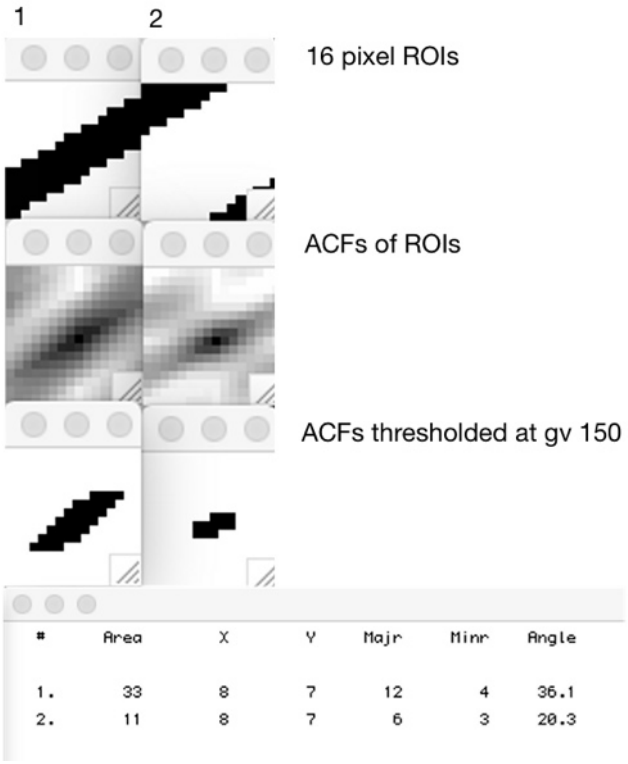


Figure 4.18: Explanation of the diversity shown in fig. 4.17: two ROIs of 16 pixels (1 and 2) were taken from different locations of the dummy image and show the variation of image shapes. The original and thresholded ACFs result in substantially different measurements for Area, Angle and Major and Minor Axes.

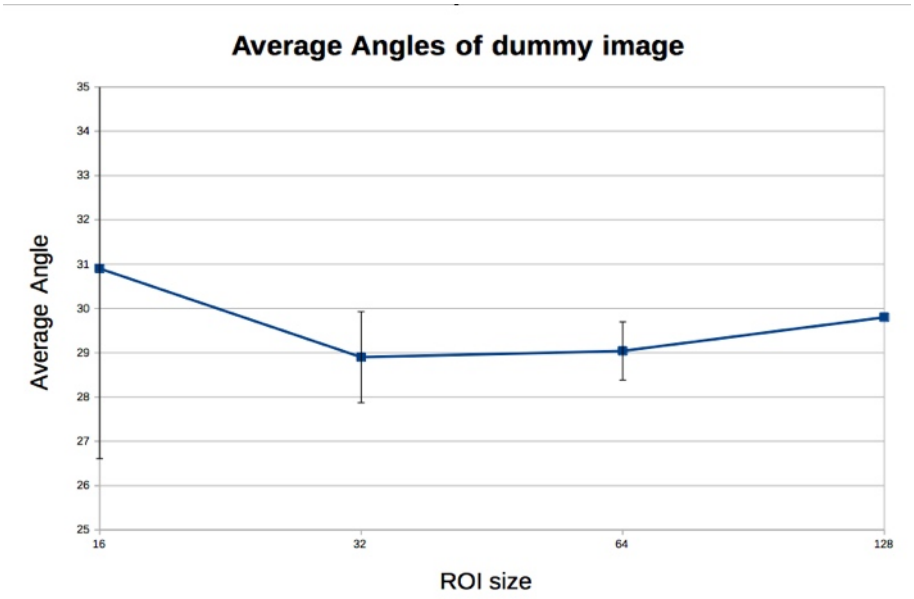


Figure 4.19: Average Angle measurements of the Linedummy image for ROI sizes 16 to 128. The image lines are angled at 30 degrees. Although the angles of the ACF particles vary at every ROI size due to ROIs covering more than one line, they do not deviate widely from 30 degree.

addition or removal of just a single pixel, and at low thresholds the central particle can merge with other, neighbouring areas of high grey value, thereby suddenly changing shape and orientation, too. Figs. 4.21 to 4.23 show the graphs for the values from 100 to 200, and for illustration the graph of *AngleR* of threshold values not included in the previous graphs is shown in fig. 4.24. The corresponding graphs for *AxanisAve* and *AxanisDev* are not shown.

It can be seen that the graphs for all thresholds between grey values 100 and 200 follow roughly the same trend. Variations are more influenced by ROI size than by threshold values. For both *AxanisAve* and *AxanisDev* the graphs seem to converge at large ROI sizes. This is unsurprising, as large ROIs sample larger areas within the same image, and are likely to be more similar than small ROIs. In these graphs there also appears to be

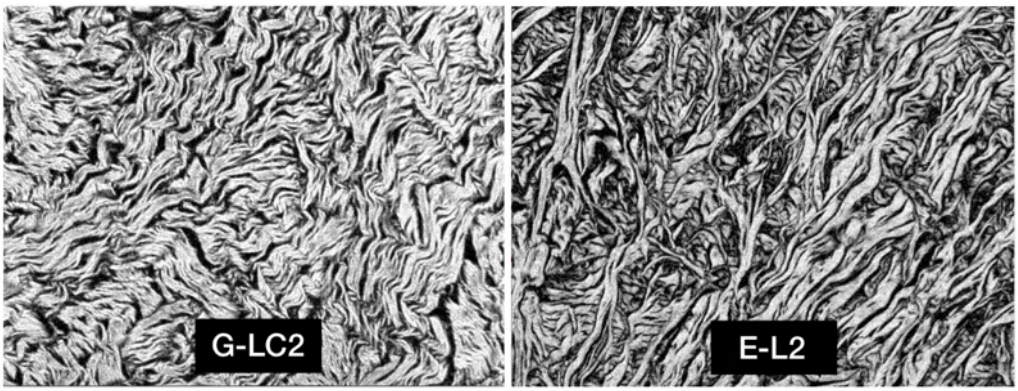


Figure 4.20: The two images that were used for testing thresholds. G-LC2 has a denser fibre structure than E-L2.

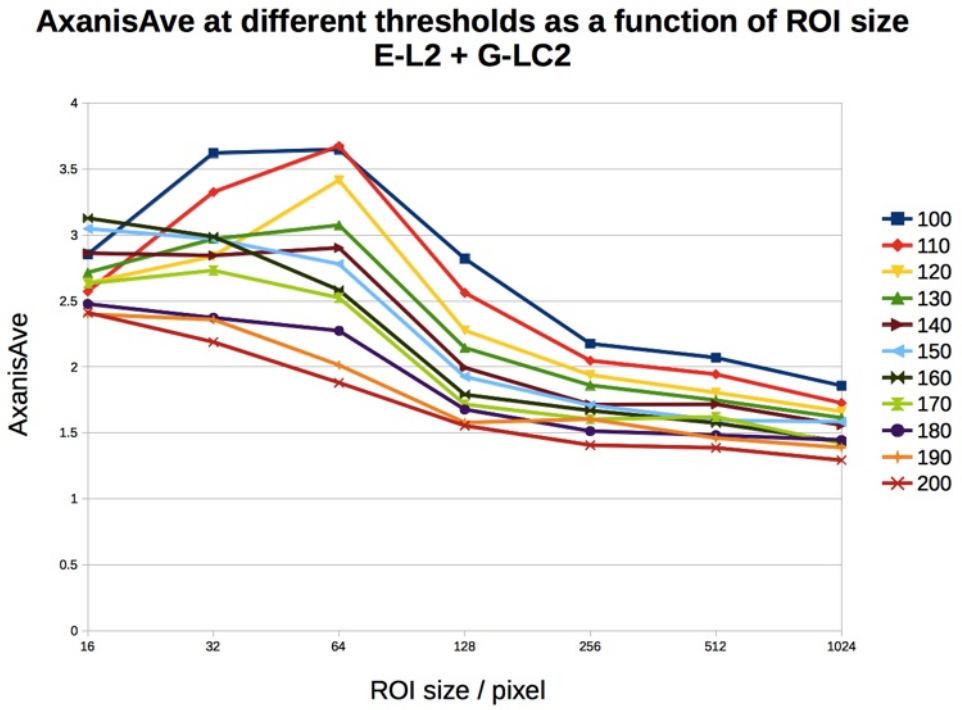


Figure 4.21: Comparison of the influence of threshold values on *AxanisAve* values. Thresholds shown from grey values 100 to 200.

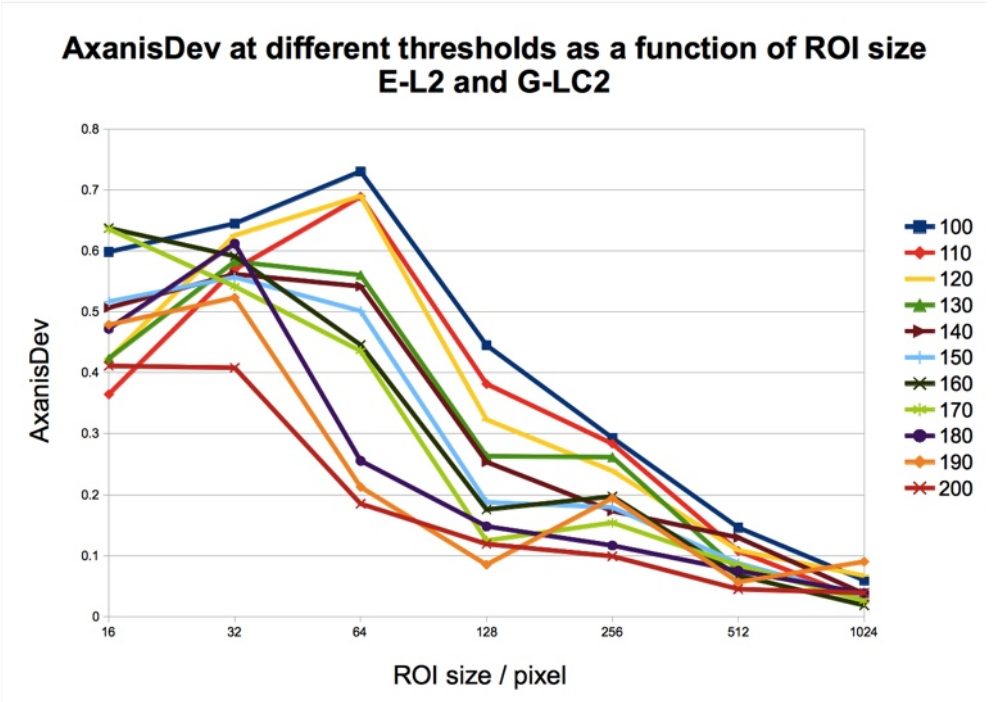


Figure 4.22: Comparison of the influence of threshold values on *AxanisDev* values. Thresholds shown from grey values 100 to 200.

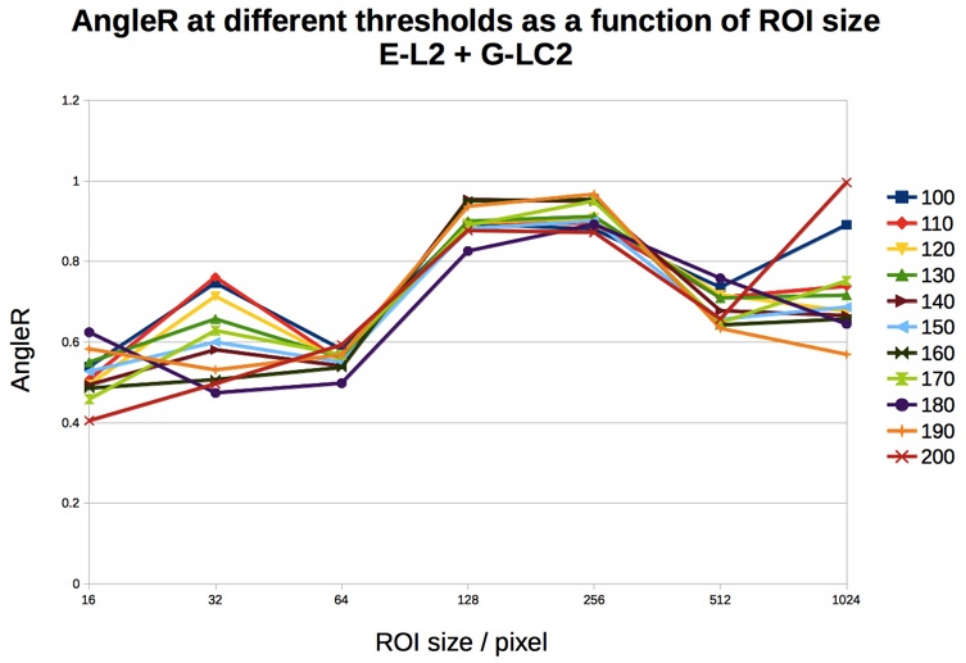


Figure 4.23: Comparison of the influence of threshold values on *AngleR* values. Thresholds shown from grey values 100 to 200.

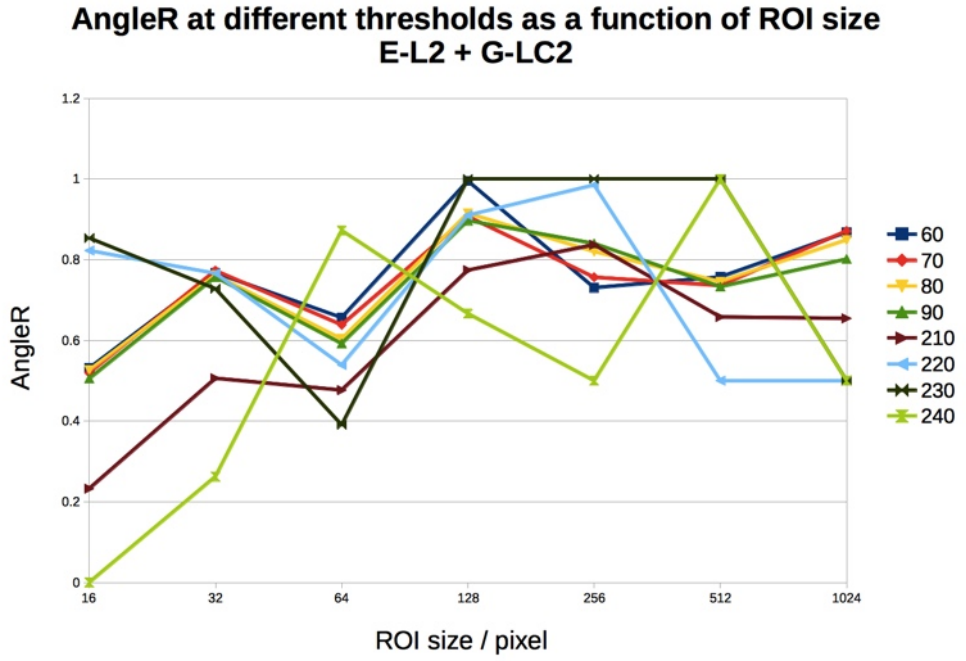


Figure 4.24: For clarity, in 4.21, 4.22 and 4.23 the thresholds shown range from grey value 100 to 200, although the thresholds tried range from 60 to 240. The threshold graphs outside the 100 - 200 range showed erratic behaviour and were therefore omitted in those figures. Here the threshold graphs excluded from the above graphs are shown for the example of AngleR. Each point in the graphs consist of three data points per image.

a slight gradient with higher values of these measures for lower threshold values and vice versa. This can be explained by the fact that at lower thresholds particles are larger and therefore will have a greater Axanis. Since the ACF analysis method has thus been shown to be relatively insensitive to the choice of threshold value within the 100 - 200 range, for further analysis of the images a grey value threshold of 150 was chosen.

4.5.3 Analysis of sample images

Different regions of the eye distinguish themselves by having different biomechanical properties mediated by variations in the organisation of their collagen fibres. Initially, the six different regions selected for comparison (fig. 4.5) were analysed in all three eyes. The intention of the analysis strategy was to generate quantities that may be used to numerically differentiate the eye regions, either overall or at distinct length scales. All images were subjected to macro analysis at ROI sizes from 16 to 256. When the images were grouped into their respective eye regions, and *AxanisAve*, *AxanisDev* and *AngleR* values were plotted against ROI size, the graphs showed no discernible differences in *AxanisAve* and *AxanisDev*, and only a small difference between C, M and the other regions in *AngleR*. (Data not shown.)

Upon visual inspection of the individual images, it became apparent that the eye region groups contained images that were very dissimilar from each other within the same group, and very similar images were found in different groups. This might account for the lack of distinctive characteristics between these groups.

In addition to examining the graphs, a statistical test was used to find similarities between the images. The two-sample Kolmogorov-Smirnov (KS) test was chosen, mainly because no previous knowledge of the nature of the distributions is necessary for its applicability. This is especially important for this analysis, as the spread of values is not primarily the result of any errors, which might suggest a Gaussian or Poisson distribution. Here, the shape of the distribution is mainly a true sample characteristic, and so no assumptions can or should be made. In the KS test, the values of the two samples to be compared are arranged in order of size and a cumulative distribution function (every next-higher value is added to the previous ones) is plotted. The cumulative distribution function of the two samples are compared and the largest distance between them is found. This is the KS value, which is then used to find the probability that the two samples are taken from the same distribution (following the calculation described in [115]. A high KS value is equivalent to a low probability (KS probability) that the null hypothesis, that the two samples are taken from the same distribution, is correct. A low KS value means that the KS probability that the two samples are taken from the same distribution is high. Thus, high KS probabilities indicate similar distributions, and low KS probabilities indicate dissimilar ones. Since the eye region groups had been found to be somewhat dissimilar by visual inspection, also individual images were compared with each other. Since *AxanisAve*, *AxanisDev* and *AngleR* are whole-image values, yielding one value per image and not a distribution of values, the distributions of related values were compared: firstly, the *Axanis* of all the ACF particles in one image at one ROI size, and secondly the

AngleDev, the deviation of the angle of each ACF particle from the average angle of the image. If two images were from the same group, it was envisaged, their distribution of the *Axanis* and *AngleDev* would be sufficiently similar to yield a high KS probability.

Although the KS test is robust in the face of different distributions, its test strength is low. The comparisons of paired images with all ROI sizes combined, as well as separated by ROI size, did not show sufficiently high correlation to be useful for image categorisation. On average, the *AngleDev* comparisons led to higher KS probabilities, but even here the highest value hardly exceeded 44% and in the case of *Axanis* the highest KS probability did not reach 35%. Furthermore, the overall distribution of KS probabilities for all pairs was very much weighted towards the low end. To illustrate this, the histograms of the KS probabilities for *Axanis* and *AngleDev* are shown in fig. 4.25.

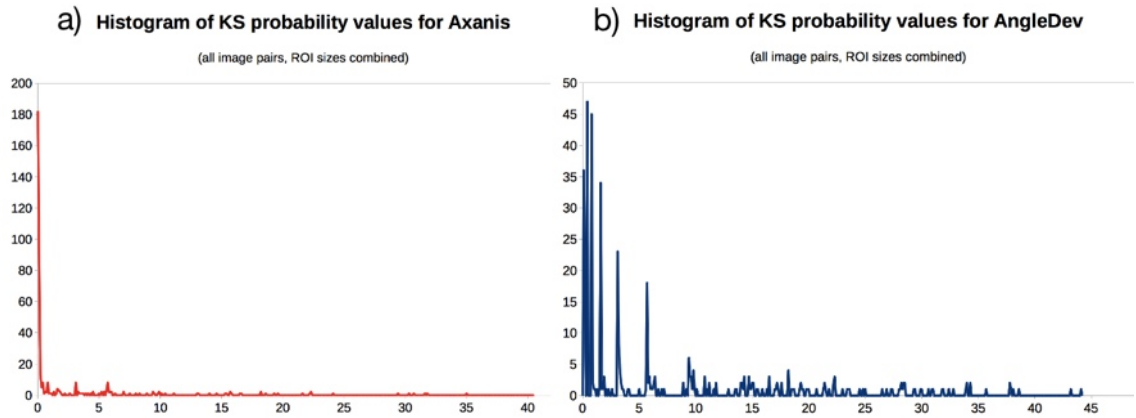


Figure 4.25: Histogram of KS probabilities for comparison of all image pairs. Each image was compared to the others at the same ROI size and the KS probabilities for each ROI size were averaged for this image pair. This averaged KS probability was entered in the histogram, a) for *Axanis*; b) for *AngleDev*. 406 data points in each case.

Thus, the KS test is unable to detect statistically significant similarities between most of the images, and is therefore not suited as a tool for meaningfully grouping similar images together into categories.

The images that were initially grouped by eye region from which they were thought to have been taken, were found by simple visual inspection to be quite divergent in appearance. However, their appearance showed obvious similarities that were not in line with the eye regions from which they were thought to have been taken. This was in contrast to previous microscopy observations, in which there was the overall impression that tissue structure was characteristic of eye region. It could not be excluded that the process of dissection, embedding and sectioning of the eyes (with its aforementioned problems) may have caused inaccurate region selection during imaging, for instance when the plane of sectioning was in fact more oblique or in a different part of the eye than expected.

Using visual appearance alone, however, most images could easily be divided into three overall groups, henceforth called A, B and C (fig. 4.26), with only two images entirely unclassifiable and four where a decision between B and C could not be clearly made. Group A images are characterised by thin parallel fibres which may have small-scale crimp but are not condensed into larger bundles or intersect, and contains only images of the cornea.

Group B is characterised by mostly, but not entirely, parallel fibres with low level of larger-scale crimp, but these are of varying thicknesses and also show some degree of fusing with and splitting from one another. Group C contains images with areas of dense fibres, which can abruptly stop or adjoin each other, and can have a wide variety in directionality and crimp. Group A contains 3 images, group B 12 images and group C 8 images.

Plotting the graphs for *AxanisAve*, *AxanisDev* and *AngleR* against ROI size for groups A, B and C shows that *AxanisAve* is higher for group B than for groups A and C, and *AngleR* is lower for group C than for groups A and B. *AxanisDev* shows very similar behaviour for all groups, with group B being slightly higher (figs. 4.27, 4.28 and 4.29)

It can be seen that a combination of the curves of *AxanisAve*, *AxanisDev* and *AngleR* can distinguish between the groups. Figs. 4.30, 4.31 and 4.32 show the same data as in figs. 4.27, 4.28 and 4.29 but plotted for each group separately.

Here it can be seen more clearly that group A has very low standard deviations overall, indicating close resemblance between the images. The *AngleR* starts at a value of just above 0.5 and rises to nearly 1, in accordance with the visual impression of crimp (i.e. more change in angles) at small scales, and largely parallel fibres (very little variation in angles) at larger scales. The *AxanisAve* stays almost at the same level through all scales, reflecting the fact that there appear to be no different features at larger scales than the ones present at small scales. The *AxanisDev*, i.e. the degree to which the *Axanis* of the ACF particles vary in an image, also stays reasonably similar for all ROI sizes, although it decreases slightly at larger scales, which is probably due to smaller scale variations being subsumed in larger ROIs.

Group B shows larger standard deviations than group A, which is expected, due to the higher variability within the image structures of this group. The *AngleR* follows a very similar pattern to group A, and when the images of these groups are compared, there is indeed a very similar pattern of directionality. At small scales group B shows a slightly larger value of *AngleR*, which may be due to the absence of crimp at this scale, which is present in group A. The *AxanisAve*, however, is significantly higher in B, which appears to be a consequence of the greater variety of shapes in these images, caused by fibres of different thicknesses crossing and merging and by larger gaps between the fibres, which are neither present in group A nor C. The *AxanisDev* is slightly larger in B than in A or C, which is most pronounced at ROI size 64. This is an indication that at this length scale there is variation in the *AxanisAve* that is not as pronounced in the other groups. Indeed, in group B, 64 pixels roughly corresponds to the thickness of the bundles and gaps between them that cannot be seen in groups A and C.

Group C shows a very similar *AxanisAve* curve as group A. (This was at first surprising. However, when the look-up table (LUT) is inverted in group A or C and the images then compared to the other uninverted group, they appear more similar to each other. The apparent difference to the eye arises from the pattern of predominantly white on black in one, versus black on white in the other. The ACF only analyses shapes and does not distinguish between dark on light or light on dark. The main difference then appears to be the variety of directions. This is another reminder that the human visual system is biased

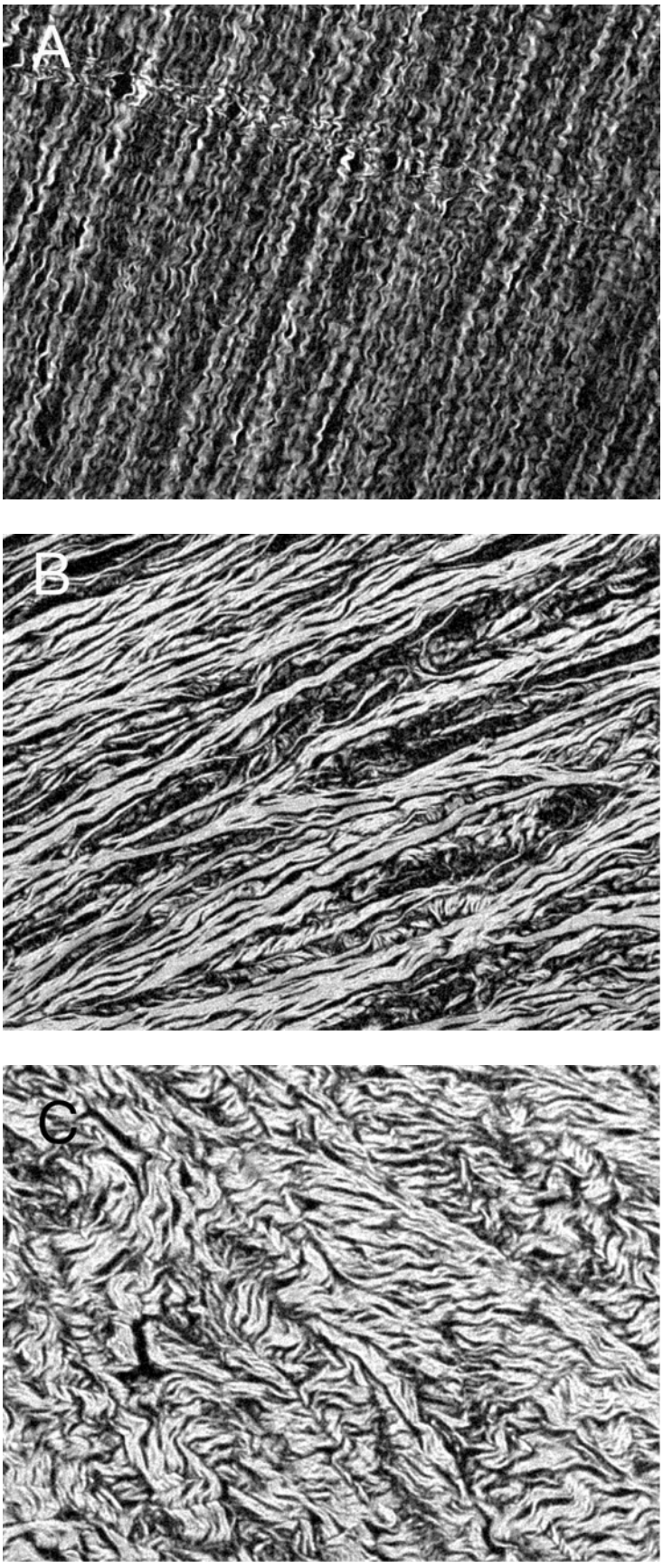


Figure 4.26: Typical examples of PLM images from group A, B and C, as classified by visual appearance. All image sizes are 1022 x 722 pixels, corresponding to 689 μm x 487 μm (A), 691 μm x 488 μm (B) and 522 μm x 390 μm (C).

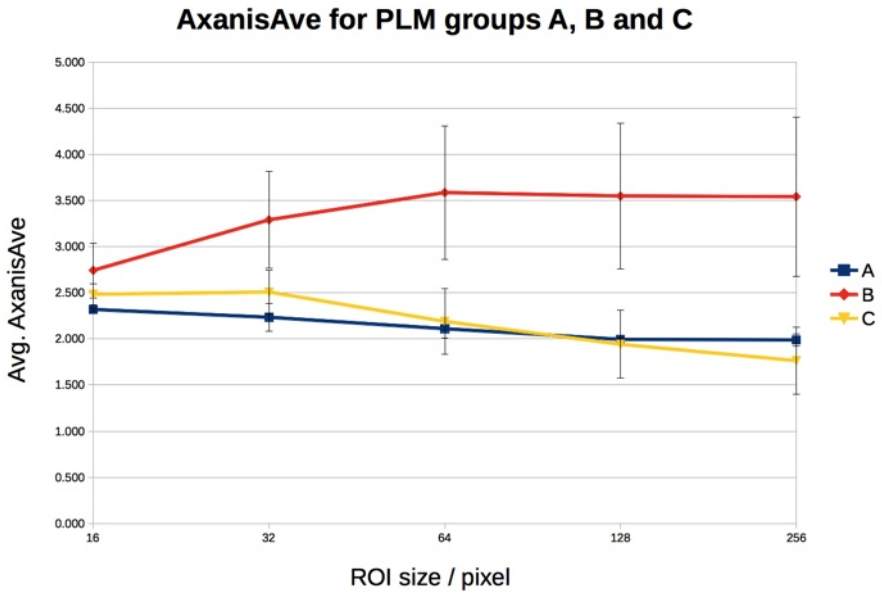


Figure 4.27: *AxanisAve* for groups A, B and C. A: n = 3, B: n = 12, C: n = 8.

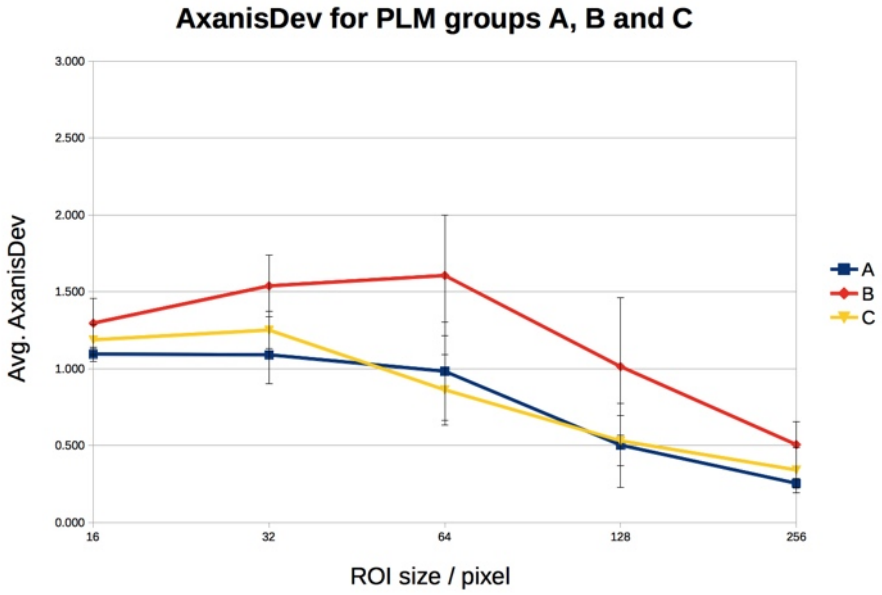


Figure 4.28: *AxanisDev* for groups A, B and C. A: n = 3, B: n = 12, C: n = 8.

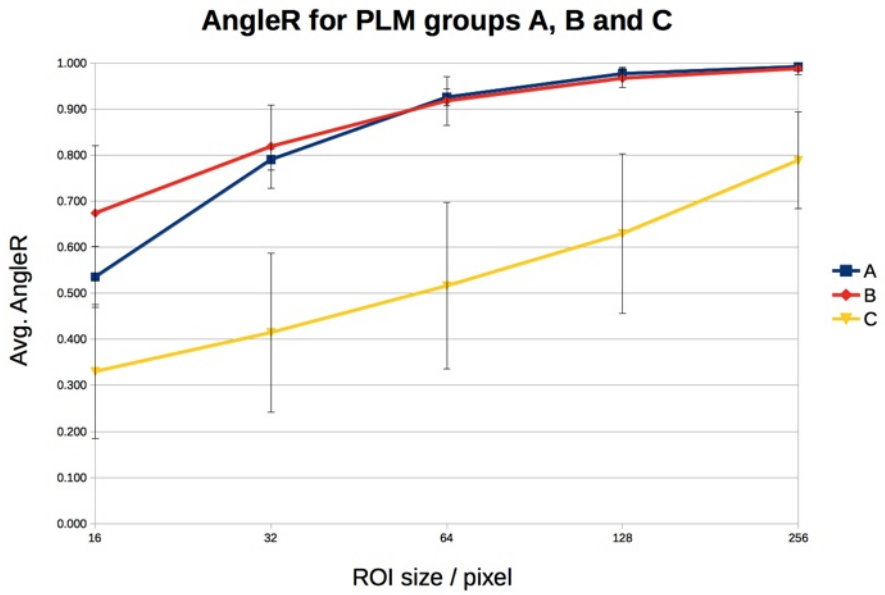


Figure 4.29: *AngleR* for groups A, B and C. A: n = 3, B: n = 12, C: n = 8.

toward different cues than those relevant to computer analysis.) The *AxanisDev* here, like the *AxanisAve*, is very similar to group A. As expected, the *AngleR* values are lower in group C than in the other groups, and show considerably larger standard deviations. This reflects the wider range of angles in these images, and perhaps also a greater variability between images.

It is now possible to compare an individual unclassified image with these groups and determine its relationship with the groups. Since the standard deviation is a useful measure of meaningful vicinity to a point, the way comparison between a single image and the groups is made is by calculating the distance of the image point to the comparison point relative to that point's standard deviation. This value s is calculated as

$$s = \left| \frac{(\text{group point value} - \text{test point value})}{\text{group point standard deviation}} \right|$$

This quantity was labelled the similarity metric, or *Simval*, and can be averaged over all ROI sizes. Different *Simvals* have to be calculated for the different metrics *AxanisAve*, *AxanisDev* and *AngleR*, and for each of the groups. Hence, for every image to be tested there will be 9 *Simvals*. An arbitrary threshold needs to be selected at which an uncategorised image is accepted as belonging to a group, or rejected. A sensible value could be a distance of one standard deviation. Thus, if the *Simval* is less than 1 for a certain metric compared to a certain group, the image is regarded as belonging to this group as far as the metric is concerned. If *Simval* is less then 1 for all metrics for a particular group, the image can be accepted into the group overall.

As more images are tested and incorporated into the groups, the collective group points and standard deviations will change, and new images will be compared to these new group values. Thus, the groups can both be honed to a narrow set of characteristics, or widened to encompass looser associations, depending on the choice and stringency of entry criteria.

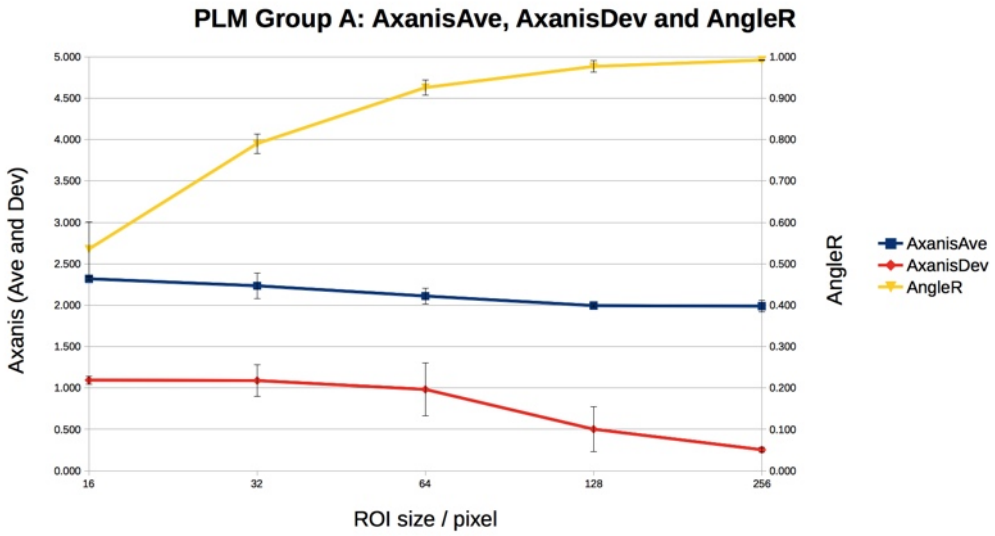


Figure 4.30: *AxanisAve*, *AxanisDev* and *AngleR* curves for group A. Note that *AxanisAve* stays constant throughout all ROI sizes, *AxanisDev* decreases only slightly, and *AngleR* rises from a medium value to almost 1 at large ROIs. The error bars are standard deviations from the variation between images, n = 3.

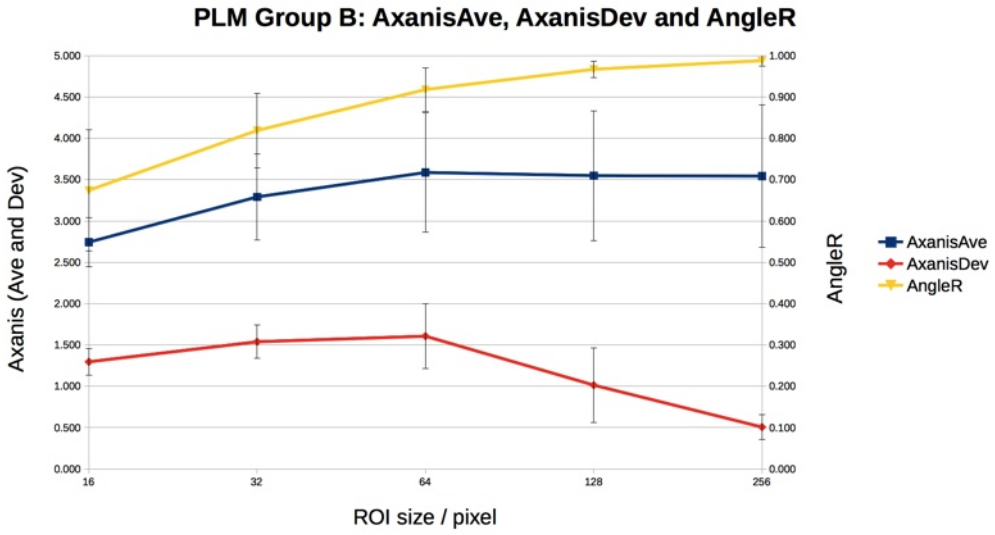


Figure 4.31: *AxanisAve*, *AxanisDev* and *AngleR* curves for group B. Here *AxanisAve* is at first similar to A, but slightly rises at larger ROI sizes, *AxanisDev* is similar to A, and *AngleR* starts at a higher value than in A, but likewise rises to almost 1 at large ROIs. The error bars are standard deviations from the variation between images, n = 12.

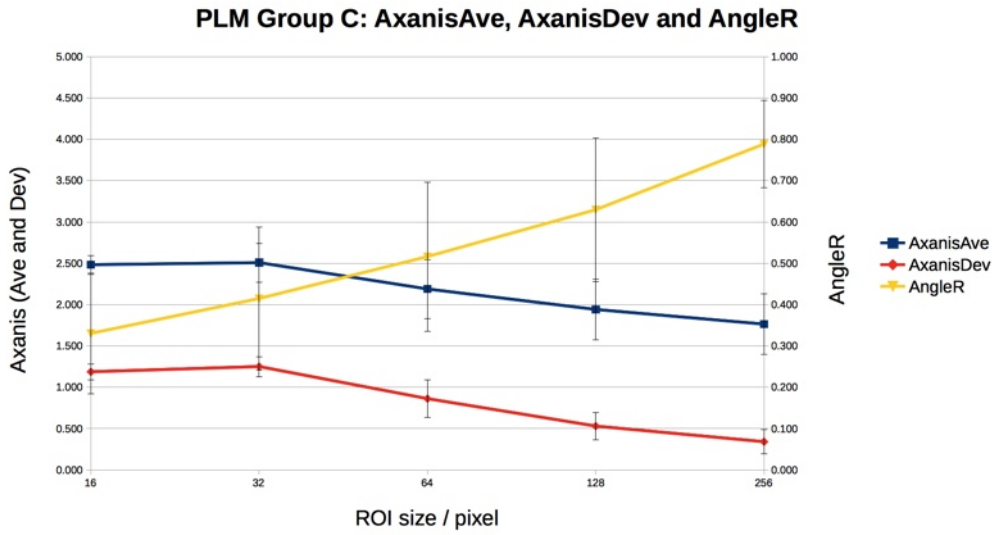


Figure 4.32: *AxanisAve*, *AxanisDev* and *AngleR* curves for group C. The *AxanisAve* behaves very similar to A, and the *AxanisDev* curve follows the *AxanisAve* trend almost exactly. The *AngleR* begins at a lower value than in A and B and rises linearly to a value of only 0.8 at large ROIs. The error bars are standard deviations from the variation between images, n = 8.

When images are compared by metric, rather than overall, new groups may emerge, which may be close to one group by one metric, but close to another group by another metric. Also, subgroups may be identifiable within the larger groups. For example, there may be clusters of images that all lie within the chosen standard deviation limit, but some of these may be just within this boundary, whereas some others may be close to the group core values. It is also conceivable that formation of such clusters can point to links between the metrics and certain image features that would not have been obvious otherwise. This analysis allows for development of categories and further analysis of data from large image sets.

As an example, an image that was not classified in the visual categorisation (GM2, fig. 4.33) was tested in this way. The image was subjected to macro analysis and a graph of its metrics was generated (fig. 4.34). Note that error bars are absent from this graph, as the data points are one value derived from the whole image at one ROI size. In fig. 4.35 this is compared with group C, as an example. The relative positions of the GM2 graph points to the error bars give a first indication of similarity. The numerical values for the comparisons with all groups are shown in table 4.2. The *Simvals* to group A are high throughout all metrics, making this image clearly dissimilar to group A. In comparison to group B, however, GM2 is similar as far as *AxanisAve* and *AxanisDev* are concerned, but dissimilar with respect to *AngleR*. The structures in the image are somewhat like group B, but the variability of angles is not. Comparing to group C, GM2 is dissimilar with respect to *AxanisAve* and *AxanisDev*, but similar in *AngleR*. Inspecting the image verifies this conclusion. The long, within some areas largely parallel fibres with little crimp are not dissimilar to structures in group B, but the whole image shows great variability in fibre direction overall, as often seen in group C images.

An image from one of the groups was also chosen and tested, and its *Simvals* were within one standard deviation from the group it was taken from for all metrics (data not shown).

Figure 4.33: Unclassified image GM2.

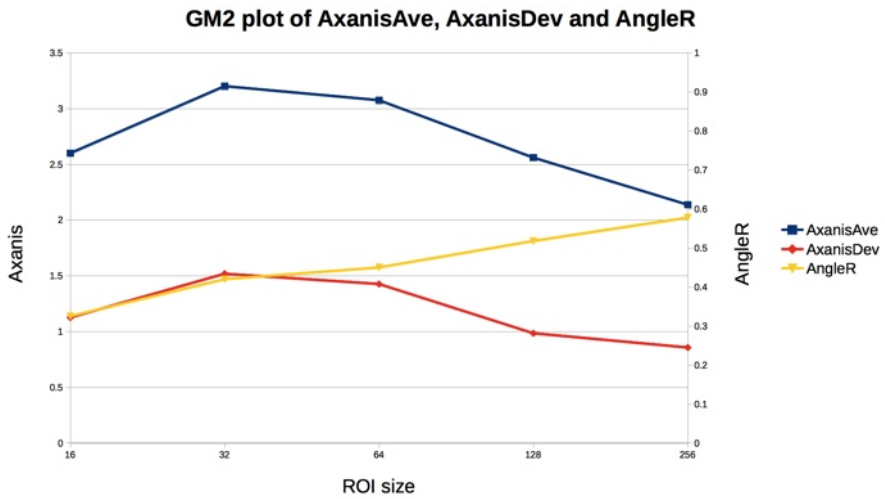


Figure 4.34: *AxanisAve*, *AxanisDev* and *AngleR* from unclassified image GM2.

This demonstrated the usefulness of this analysis strategy for images of the type we analysed, and shows that there is potential to develop the method further to adapt them to other image types, and situations where a similar kind of textural data may be interesting.

4.6 Future work

As in the case of the AFM analysis, writing this approach into Pascal and incorporating it into Image SXM, or converting it into an ImageJ plugin in Java, will be the next

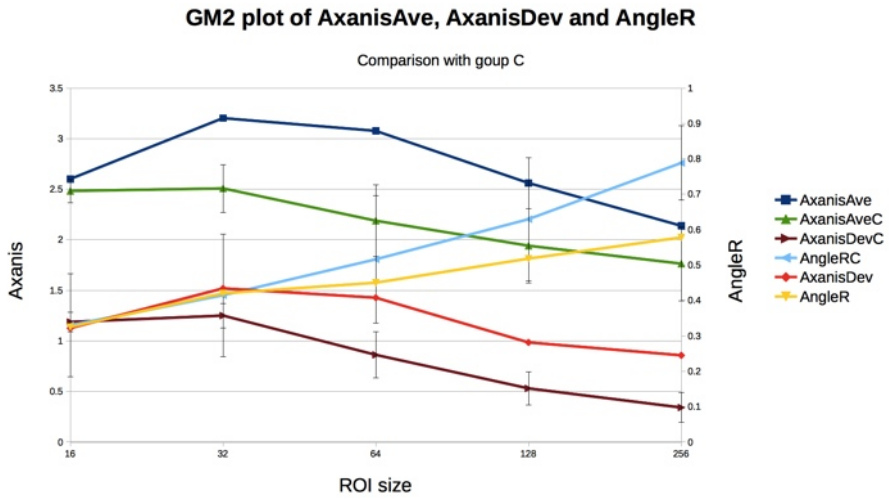


Figure 4.35: *AxanisAve*, *AxanisDev* and *AngleR* from unclassified image GM2, compared to group C. Note that the corresponding graphs for *AxanisAve* (blue and green) follow similar trends but are further apart than one standard deviation (error bar). The situation is similar for *AxanisDev* (red and brown). In the case of *AngleR* (yellow and light blue), however, the yellow graph for GM2 lies mostly within one standard deviation from the light blue curve of group C.

GM2	A	B	C
AxanisAve	15.333	0.847	1.843
AxanisDev	5.113	0.802	2.324
AngleR	158.740	13.585	0.615

Table 4.2: Comparison of unclassified image GM2 with groups A, B and C for all metrics (*Simvals*) in units of standard deviations.

step to make this method accessible to a wider range of users. Whereas the analysis of each image by macro took between two and three hours, removing the additional step of compiling the macro language will shorten this period significantly.

An important step in improving the method is adapting the threshold. Currently, a carefully selected, but nevertheless constant threshold is used. This has several disadvantages. Firstly, some ACF particles become meaninglessly small, whereas others become so large that they would spill over the ROI boundaries, thus falsifying the measurements. Due to the large number of ROIs per image, this is not a major problem, as this affects a relatively small number of particles. A better strategy, nevertheless, would be to have a set of criteria for thresholding each ACF. For example, increasing the threshold up to a value where the major axis of the particle is less than the ROI size (i.e. it is not artificially curtailed by the ROI size), but only threshold here, if at a higher threshold the area of the central particle does not drop in sudden jumps (this would be the case if the central particle was in fact the result of a fusion of neighbouring grey value peaks, which would separate at higher thresholds), in which case threshold at this higher value after the jump has occurred. Since these measurements can only be made after the ROI has been thresholded and the Particle Analysis carried out, in order to implement such an adjustment, for every single ROI taken from the image a number of thresholds would have to be applied, a Particle Analysis carried out for each, the measured values compared with previous ones and the thresholding undone again ready for the next attempt. Not only would a considerable number of calculations have to be made, which would make the analysis prohibitively time-consuming, but also a number of user arrays would be needed that exceed the number available to the macro user. It is conceivable that other threshold adjustments might be preferable for different images, but details about these images would be needed to assess the optimal strategy.

Another analysis route derived from this strategy might be to compare the relative importance of ROI sizes for the identification of image types. Since ROI sizes relate to length scales in images, the ROI sizes most meaningful for identification of groups will be slightly larger than the image features in which the groups differ. This could be a valuable pointer for finding different image characteristics for distinguishing groups, which might not be obvious to visual examination.

The most obvious future direction, however, is to use a large group of images similar to the small number analysed here and carry out a categorisation analysis on this group. While the major considerations for such a project have been outlined here, carrying out the analysis is apt to lead to further insights and deeper knowledge of the procedure.

4.7 Conclusion

An algorithm was developed for the analysis of complex, information-rich images of collagen obtained from polarised light microscopy. The analysis uses three novel, semi-textural metrics, applied to different length scales, that capture overall image qualities rather than individual fibre characteristics. The use and successful quantification of these metrics allows categorisation of images into groups of similar appearance, independent of a priori

knowledge of groupings. With large numbers of images, the analysis is capable of making clusters and substructure in the groupings evident. Furthermore, the likelihood, based on these metrics, for an individual unclassified image to belong to a previously established group can be numerically evaluated. Since no user intervention is necessary in this analysis, the throughput is limited only by availability of data and computing power. This prepares the way for quick and efficient analysis of large volumes of digital data that modern microscopy techniques are able to produce.

The principle of the analysis can be extended to a greater number of metrics, and in principle also to different metrics suitable to other image types.

Chapter 5

Second Harmonic Generation

In this chapter a different microscopic technique, Second Harmonic Generation (SHG) microscopy, is used to take images from the same slides that were imaged by Polarised Light Microscopy (PLM) in the previous chapter.

Although PLM, in this context, and SHG are both specific for collagen (although different in their principles of detection, see below), their practical scopes differ markedly. The information that can be gleaned from PLM is relatively limited, showing structures in a two-dimensional plane, at a relatively low resolving power, and only after chemical fixation of the tissue sample. SHG, on the other hand, does not require any tissue fixation and can in principle be used on live samples without causing damage. It has the ability to generate a signal from a tissue depth of up to 200 μm ([116]) with confocal precision, allowing investigation of three-dimensional structures, and resolving power can with modifications be extended beyond the Abbe limit (for example [117]). Furthermore, SHG generates photons in tissues in both forward and backward scattered geometries, and this makes it possible to design instruments that rely on backscattered photons only, which can gather information by simply holding the instrument close to a living sample (for example [118]). On the other hand, the cost of PLM is low enough for very widespread use to be feasible, whereas the cost of an SHG microscope is prohibitive for any but specialised laboratory setups. In this study no use was made of the superior abilities of SHG microscopy, due in part to constraints in instrument availability, and in part because this was not necessary for the purpose of comparing macro performance.

This second analysis was done to compare the performance of the macro algorithm when applied to images derived from a different (but in scale comparable) imaging modality. Therefore, the same macro analysis was performed on these images, and the results of the analysis of the SHG images was compared with the results from the PLM analysis discussed in the last chapter. In keeping with the overall structure of the thesis, the analyses of these two microscopies are kept in separate chapters, even though the chapters are closely related.

5.1 Principles of Second Harmonic Generation Imaging

5.1.1 Second Harmonic Generation

When photons interact with matter, a large range of interactions can and do occur. Most involve initial absorption of photons followed by subsequent emission of photons. Describ-

ing these interactions reveals that in some cases the relationship between the incident photons and the emitted ones is linear (linear effects, such as scattering or reflection), whereas in others the interaction is described by a higher order term (non-linear effects). The non-linear effects become more dominant at higher light intensities. An example of such a non-linear effect is Second Harmonic Generation (SHG), in which two photons of a certain energy are absorbed, and a single photon of twice the energy of the absorbed photons is emitted (fig. 5.1).

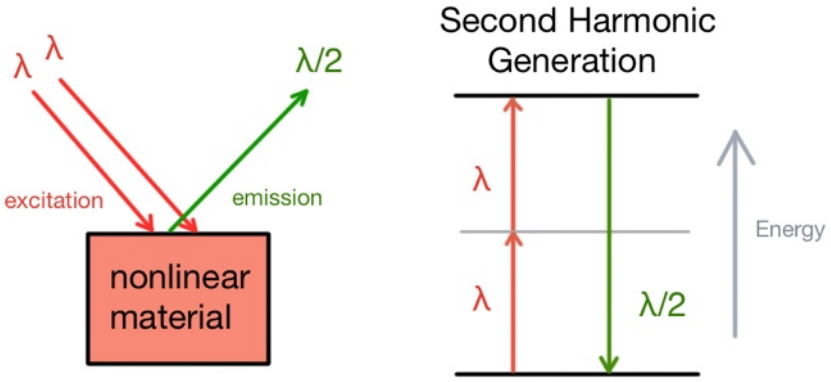


Figure 5.1: Schematic representation of the generation of second harmonic photons. Excitation light hits the target material, which converts two excitation photons of wavelength λ into one emission photon of wavelength exactly $\lambda/2$.

This effect only becomes noticeable at high light intensities and is exhibited only in crystalline substances which are non-centrosymmetric on the size scale of the wavelength of light. In such materials, when the polarisation of the valence electrons induced by the intense electromagnetic field of the incident light becomes saturated, non-linear effects like SHG appear, and increase with the intensity of the field. Collagen is recognised as a non-centrosymmetric crystal. The crystallinity state of collagen is complex, however. The diameter of a collagen type I fibre (composed of multiple fibrils, each containing numerous collagen molecules) ranges typically between 50 to 500 nm, which corresponds well to the wavelength of visible light. The amino acid sequence is highly repetitive and aligned. But X-ray diffraction studies show significant amounts of disorder, and collagen has been described as a liquid crystal with dynamic properties ([53]), which refers to the side-by-side alignment of molecules and fibrils ([119], [120]). The gap region where the 300 nm long collagen molecules join head-to-tail is thought to contribute significantly to this disorder along the direction of the long fibre axis ([121]). Hence, there is pronounced symmetry breaking both along the long axis and across the diameter of the fibres. In addition, the interface of the collagen fibrils, fibres, fascicle and tendons with their surrounding medium is another location where symmetry breaking occurs. It is not yet clear where exactly in the collagen structure the second harmonic signal arises. Since collagen is highly varied in its structure and composition, dependent on tissue, age, use and health state, etc., it is likely that the relative contributions may vary in different situations. For review see [122], [123] and [124].

SHG is a coherent effect, which refers to the existence of a specific emission pattern of forward and backward propagation of the signal. The coherence length L_c, over which the excitation and emission signal could constructively or destructively interfere, is defined as 2π / Δk, where Δk is the phase mismatch. The phase mismatch is defined by $\Delta k = k_{2\omega} - 2k_{\omega}$, where $k_{2\omega}$ is the wave vector for the SHG photons, and k_{ω} the wave vector for the excitation photons (wave vectors are $2\pi/\lambda$). If L_c is at a minimum, then the phases of the excitation and SHG signal photons perfectly coincide. In this case, all the SHG emission is forward propagating, with respect to the excitation beam, and this condition is called perfect phase-matching. In real biological tissues, this condition is only ever approximated. Due to the fluid organisation of biological molecules, with a random component, Δk is seldom, if ever, zero in tissues, and consequently the SHG signal there consists of a forward and a backward propagating component. Under such conditions, the SHG is regarded as quasi-coherent ([125]). Phase-matching conditions can best be approximated when the SHG signal comes from a very small volume (such as a well-focussed light cone) or a very thin section of tissue. Fig. 5.2 shows an intuitive way to visualise this (adapted from [126]).

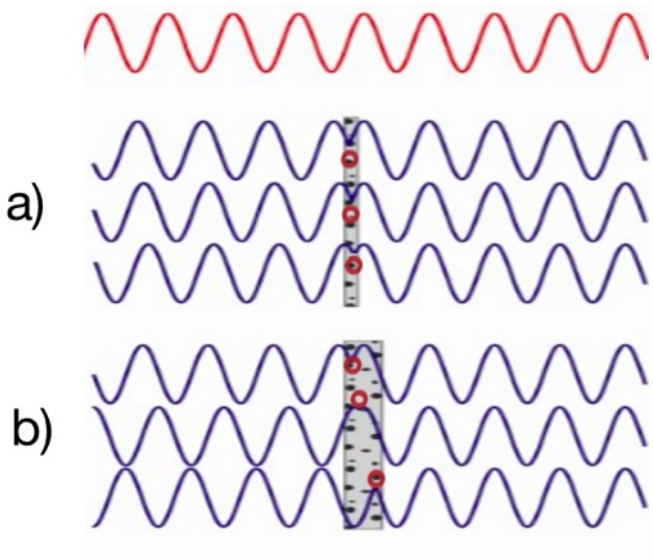


Figure 5.2: Phasematching of the SHG signal with the excitation beam (red wave): a) the source of the second harmonic generation in a sample thinner than the wavelength of light, Δk is close to zero, forward propagated waves are in phase with the excitation waves; b) second harmonics generated in a thicker sample (on the scale of the wavelength of light and thicker), Δk is larger than zero, the backward propagated waves do not have well-matching phases and therefore destructive interference can destroy some of the signal. Red dots represent harmonophores. (Adapted from [126].)

The ratio of forward to backward propagated SHG signal (F/B) in collagen is influenced mostly by the fibre alignment and direction relative to the excitation beam and polarisation axis ([127]), but other factors influence this ratio, too, such as the hydration state, age of collagen, or tissue environment. The actually measured ratio of F/B is a convolution of the forward and backward propagating signal generated at the harmonophore[1] with the

[1] A harmonophore is the smallest source unit of the SHG signal in the tissue.

secondary scattering that occurs in the tissue before the photon reaches the detectors, and is thus influenced also by the nature of the tissue environment.

In addition to the impact of the orientation of the collagen fibres relative to the excitation beam in the x-y plane, also the angle with which the collagen fibres emerge out of the horizontal influences the efficiency of second harmonic generation ([128]). This property can be used to glean additional information about tissue architecture.

The backward propagation of the SHG signal in principle allows imaging of superficial tissue *in vivo*, by eliminating the need to place the tissue between the light source and a detector, and thus shows promise for diagnosis of skin diseases non-invasively and quickly. As noted before, many diseases are linked with faulty production or assembly of collagen, and SHG provides a powerful tool for characterising the specific nature of the changes in tissues and for diagnosis (see for example [129]).

In this thesis, neither the F/B ratio, nor the sensitivity to angle in the z plane, nor the confocal capability of SHG microscopy have been utilised. This is mainly due to the nature of the samples at our disposal, and the limited time available on the SHG microscope. The PLM samples' section thickness was on the order of the size of the z resolution of the microscope (Dr Chris Thrasivoulou, personal communication), and since the purpose of the images was the comparison of the macro analysis, the F/B ratio was not of interest. The sensitivity of the signal to angle in the z plane is negligible at such thin sections ([128]). Also the direction of the fibres in the x-y plane relative to the polarisation plane of the excitation beam did not have to be accounted for, as the polarisation beam was circularly polarised.

5.1.2 The SHG Microscope

The phenomenon of SHG was first observed by Peter Franken in 1961 [130] in a quartz crystal, shortly after the invention of the laser by Theodore Maiman in 1960, which provided for the first time sufficiently intense light for SHG to occur at a noticeable rate. In 1974, SHG was first used in a microscopy context [131], albeit using whole-field illumination, which showed SHG only in very effective second harmonic generators with a high degree of hyperpolarisability of their electron shell. The restriction to very effective SHG generators was due to the low light intensities in whole-field illumination. In their study, Hellwarth and Christensen [131] imaged a potassium deuterium hydrogen phosphate crystal. A scanning approach to SHG imaging provided an improvement, which was pioneered by Gannaway and Shepperd in 1978 ([132]), in which all light could be concentrated on a small image area, thus improving the light intensity and efficiency of second harmonic generation. SHG was first used in a biological context in 1986 by Freund and colleagues [133], studying the polarity of collagen in rat tail tendon. Since then, many more studies have followed, and SHG is continuing to attract increasing interest. This is due to its ability to detect and image collagen and actomyosin complexes in tissues very specifically, and without the need for any staining procedure, both *in vivo* and *in vitro*, and to its capacity to act as a confocal microscope almost solely by virtue of the steep slope with which the SHG signal fades outside a small volume of the focal cone. No signal is generated in other

parts of the sample. SHG microscopes have been developed into off-the-shelf facilities, and are therefore now accessible to a larger number of researchers.

Fig. 5.3 shows a schematic representation of a typical SHG microscope setup, with detectors for forward and backward propagating signal and the filters that allow the excitation beam to be polarised in a variety of ways.

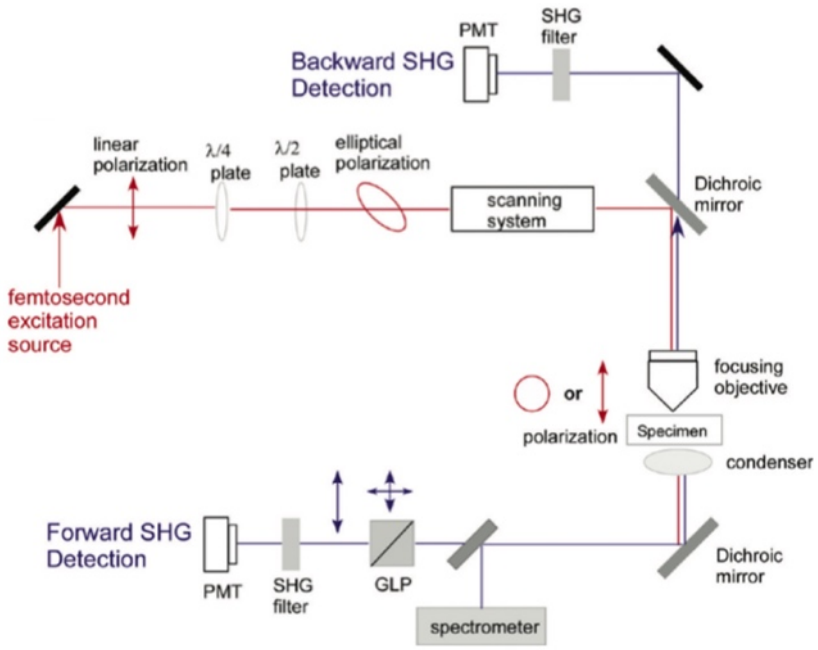


Figure 5.3: Schematic representation of an SHG microscope. (PMT = photomultiplier tube, GLP = Glan laser polariser) (Adapted from [129].)

In modern SHG microscopes for biological applications, the use of pulsed lasers allows a very high light intensity at each point in the image (in the region of 10^8 Wcm^{-2}) without destroying the sample. A modern femtosecond laser with a repetition rate of 80 - 100 MHz has a pulse duration of around 80 - 100 fs and an off-time of around 10 - 12.5 ns, allowing enough time for the energy in the sample to dissipate before the next pulse ([134]). This makes it possible to achieve good signal contrast even of small concentration of harmonophores.

5.1.3 SHG of Collagen

As already discussed, collagen is the most abundant protein in animal bodies and problems with its production, regulation, structure and deposition lead to a variety of diseases. Visualising biological structures by microscopic techniques has historically required alteration of the material with biochemical modifications, however this can lead to artifacts. SHG can visualise collagen without the need for any prior tissue preparation and with excellent specificity. Therefore, SHG is of great interest to the study of collagen and many collagen-related studies using SHG have been published. Most are from a biomedical background and try to distinguish between healthy and unhealthy collagen in wound healing and many disease states. While some use only qualitative assessment, others attempt to

find quantitative measures.

Briefly, the simplest form of analysis (applicable to any image) consists of measuring pixel characteristics of the whole image, such as average and spread of intensities or shape and features of their histogram. While these measures provide some information about the image, they do not usually provide clinically very useful information.

The analysis of the ratio of forward and backward propagating signal is attracting much interest, but the interpretation of this is not straightforward, requires confirmation by other methods, and is still subject to debate ([135]).

Nevertheless, as mentioned above, the backward propagating signal is especially useful for *in vivo* investigations in a clinical context. The development of instruments which are small and cheap enough to be used at the bed side or endoscopically is under way, and such non-invasive microscopy will rely primarily on the backward propagating signal and its quick and reliable analysis.

The use of a linearly polarised excitation beam is used to gain information of fibre directionality, since the signal intensity is a function of laser polarisation relative to fibre orientation. Taking images while rotating the polarisation axis through 180 degrees in the same field of view yields information of main directions through changes in brightness.

Fourier transform methods have also been used to extract texture characteristics from SHG images of collagen, and are most useful when combined with the use of polarisation or F/B ratio microscopy. (For example see [136].)

Texture analysis methods, such as the grey level co-occurrence matrix (GLCM), have also been used with some success. Briefly, the GLCM calculates the frequency with which pixels of the same grey value occur in a specific spatial relationship in a given image. A number of statistical measures can then be extracted from this matrix, which give textural information about the image.

Due to the limited scope of this chapter, the large body of publications concerning SHG imaging of collagen and its analysis is not reviewed here. Mostaco-Guidolin et al. [137] provide an excellent, concise review of the current state of the field, with many further references to past work.

5.2 SHG Images and Analysis Strategy

5.2.1 SHG Samples

The same set of histological slides that were used for PLM analysis was also used for SHG microscopy. (See fig. 4.5 in the PLM Chapter.) Briefly again, three porcine eyes were dissected in a medial plane with respect to the eye ball, fixed, sectioned at 4μm thickness, stained with H&E (haematoxylin and eosin, a standard histology stain that shows nuclei through haematoxylin and other generic tissue structures through eosin) and picrosirius red (PSR) and embedded in a polystyrene mountant. Neither the PSR or H&E stain, nor the mountant interfere with second harmonic generation or detection (Dr Chris Thrasivoulou, personal communication). The samples were taken to the Confocal Imaging Facility at University College London, where they were used to collect SHG microscopic

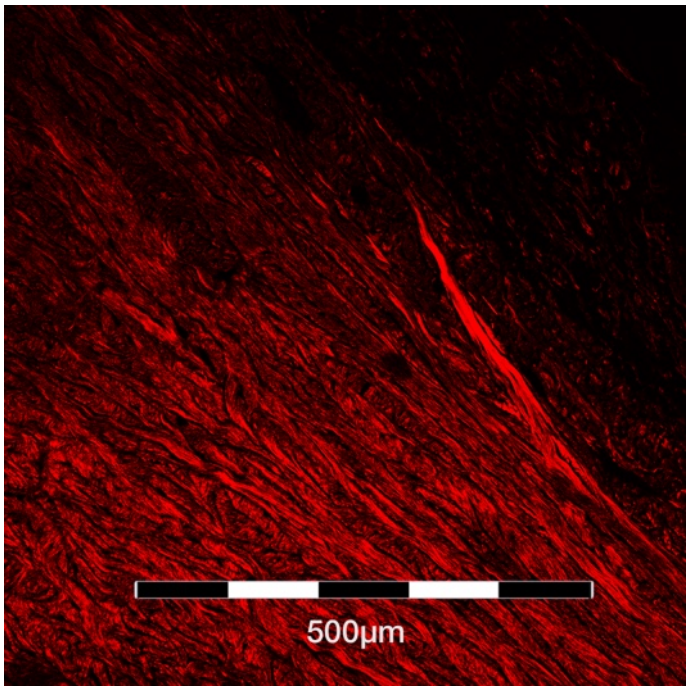


Figure 5.4: Example of SHG image at 10x lens magnification. The scalebar is shown here for information, and the scale is the same for all SHG images.

images of the same regions of each eye, but not of the identical locations as those imaged by PLM.

5.2.2 SHG Images

The SHG images were taken, with the kind help of Dr Chris Thrasivoulou, on a Leica SP8 microscope equipped with a coherent Ti-Sapphire two-photon laser with a tuneable output wavelength range from 680nm to 1060nm and a power output of 2430mW, and a Leica HyD detector. In keeping with the PLM data, a 10x objective was used. The field of view span 737μm x 737μm, corresponding to a digital image size of 1024 x 1024 pixels. These images were cropped to 1022 x 772 pixels, in keeping with the size of the PLM images. An uncropped image with scalebar is shown in fig. 5.4. The z resolution at this magnification of the instrument is 3.5μm, roughly the thickness of the tissue. Therefore, the ability of the SHG microscope to resolve deeper structures was not utilised in this case. Furthermore, quarterwave plates were positioned in the light path before and after the sample, so that the polarisation plane of the laser relative to the direction of the collagen fibres in the tissue did not need to be taken into consideration.

The SHG signal was collected in both forward and backward scattered directions. The forward scattered photons were represented in the (false colour) red channel, whereas the backward scattered photons were represented in the green channel. Since the green channel did not show any additional structures when red and green images were overlapped, and the intensity of the red channel was significantly higher, only the forward scattered images of the red channel were used for analysis.

In the PLM analysis it became apparent that the regions of the eye were too inhomoge-

neous a set of groups, and there were as many differences within the groups as there were similarities between them, for the analysis of group differences to be meaningful. Hence, images were grouped according to visual similarities and an analysis developed that could distinguish these visual groups by numerical characteristics.

For illustration, typical images representing the visual groups, A, B and C, are shown in fig. 5.5. In comparison to the PLM images these images show a greater resolution of individual fibres. Where in PLM images smaller fibres seem to coalesce into thicker bundles, in SHG the individual thin fibres remain more separate from their neighbours. It is expected that this will lead to a difference in the macro results especially at lower ROI sizes.

The images were grouped into the visual categories in the same way as for PLM analysis, and all but one of the images were assigned to the same group as their PLM counterpart. Since the attempts to image the same areas of the eye balls likely led to some spread of exact location within the regions, it is likely that the inhomogeneity of the eye region groups is a consequence of dissection, embedding and sectioning, so that tissues were presented at different angles, rather than choice of region for microscopy.

To illustrate the visual differences between both microscopies, fig. 5.6 shows images taken of equivalent regions in PLM and SHG microscopies. In the SHG image smaller fibres appear resolved than in the PLM image.

5.2.3 Thresholding

Due to the presence of thinner fibres in the SHG images, the threshold of 150 that was used for the ACF particles in the PLM analysis was not suitable, since at that threshold many particles were removed. In the PLM analysis, when testing threshold values for the ACF particles, three ROIs were selected at each size and thresholds tested in the range from grey values 60 to 240. Figs 4.21 to 4.23 show that when each ROI size is considered separately, the graph points for *AxanisAve*, *AxanisDev* and *AngleR* resemble a straight line as long as the threshold is in a useful range, but this pattern disintegrates for extreme thresholds (fig. 4.24). Therefore, a simplified procedure for adjusting the threshold was used for the SHG analysis. Only one ROI of size 128 from one image was chosen and thresholds from 10 to 240 were tested. The *Axanis* and *Angle* were plotted against threshold values (fig. 5.7), and a threshold value in the linear range of the graph was chosen for analysis of the SHG images. Since the linear range of the graphs for both *Axanis* and *Angle* is found between grey values 40 and 120, a midpoint of 80 was used as the threshold for SHG analysis.

As already stated in the previous chapter, a constant threshold is a compromise necessitated by the restriction of a macro. A threshold value that is automatically responsive to the nature of the image might be desirable, but the duration of the analysis, as well as access to internal software arrays makes this impossible to implement in a macro.

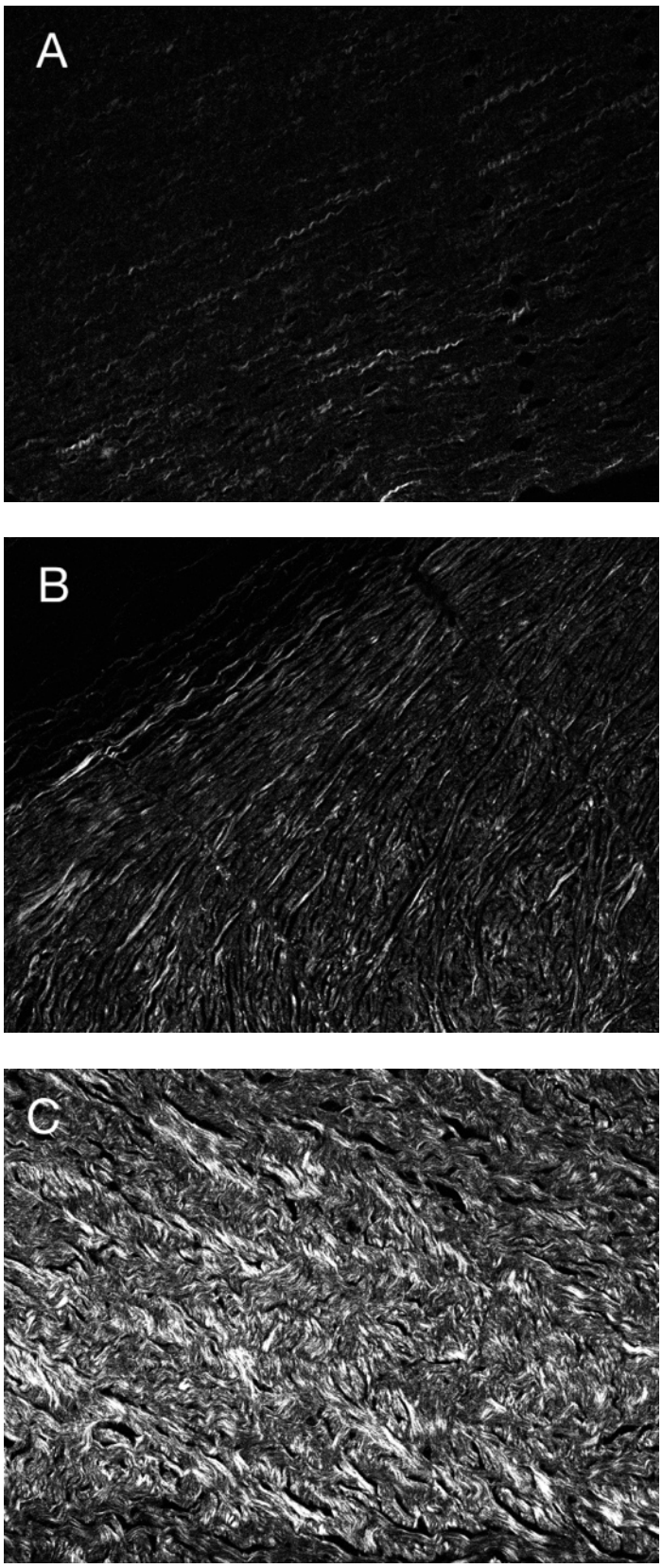


Figure 5.5: Typical examples of SHG images from group A, B and C, as classified by visual appearance. These images are the SHG equivalents of the PLM images shown in fig. 4.26. It can be seen that, while the SHG images show more detailed structure at small length scales, there is considerable similarity to the PLM images. Image size is 1022 x 722 pixels, corresponding to 739 μm x 522 μm

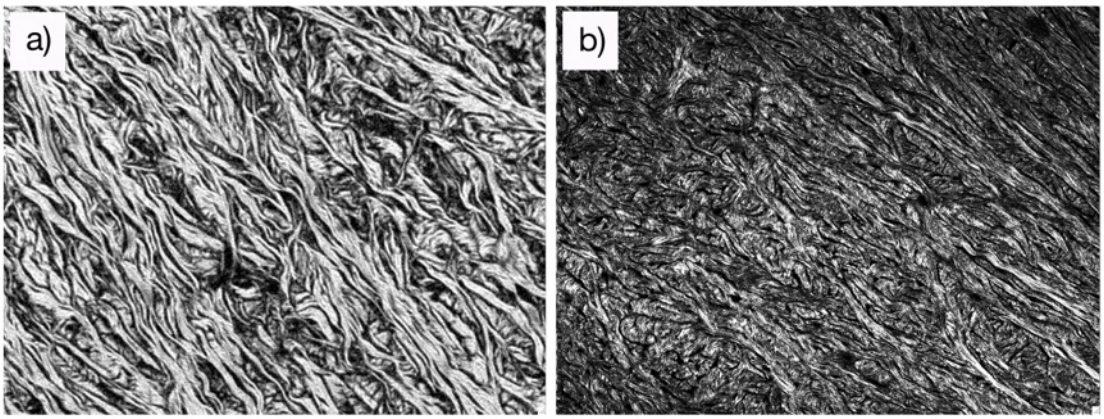


Figure 5.6: PLM (a) and SHG (b) images of similar eye regions and both in visual group B. In the SHG image smaller fibres often make up the larger ones, whereas in PLM the larger fibres are a single block.

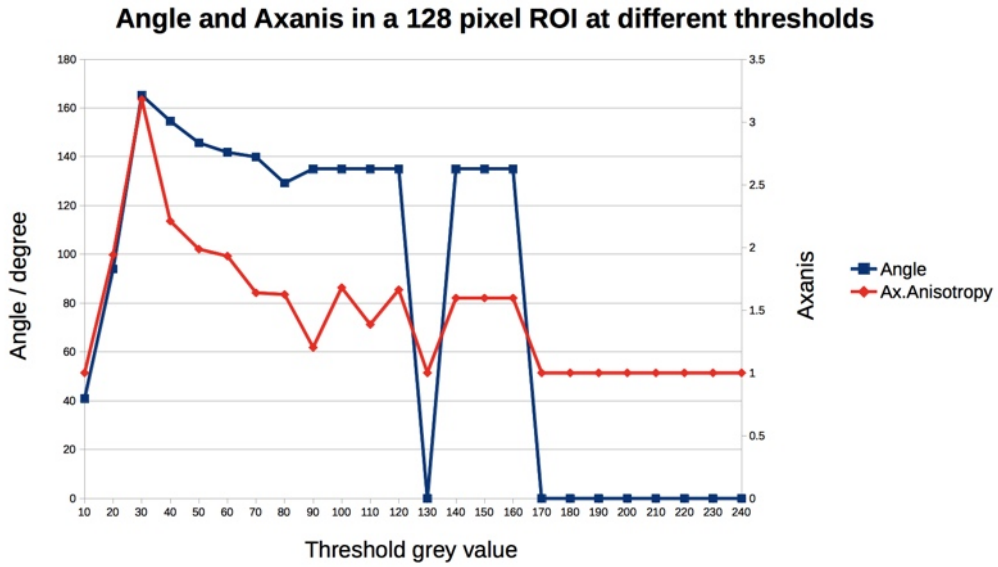


Figure 5.7: Threshold range for analysis of SHG images. The values at grey value 130 and at grey values greater than 160 meet the exclusion criteria of an *Angle* of 0, and *Axanis* of 1 (round) and a pixel area of less than 10, and would therefore be excluded from analysis of the macro data. It can be seen that the thresholds that yield meaningful data range from about 40 to 120. A threshold of 80 was chosen for the analysis.

5.2.4 Analysis Strategy

The analysis strategy for the SHG images is the same as for the PLM analysis. The rationale behind this approach is that comparing the results of both analyses can provide a sort of control of the robustness of the macro, as well as potentially allowing inferences about the information provided by different microscopies.

The characteristics of the underlying biological structures in the images can be assumed to be the same. The images are not identical, however, because the fields of view chosen are not the same. The analysis strategy aims to identify properties characteristic of a type of tissue, whether this type is initially characterised by its location in the eye or by seen similarities. These characteristics need to vary with tissue type, but not with accidental differences such as those arising from change of field of view. If an analysis was sensitive to the field of view, the analysis would not be useful.

Any fundamental differences in such defining characteristics that may arise between the PLM and SHG images themselves will be derived from the idiosyncrasies of the microscopic techniques employed, if the macro works well. As both techniques make use of different physical mechanisms to identify collagen and have different ways of forming an image, differences in the outcome are intrinsic. It might be hoped that these only lead to the kind of accidental differences that would not interfere with tissue typing, but whether this is the case is as much dependent on the correct choice of characteristic as it is on the sensitivity of the macro to circumstantial image properties.

In addition to the differences introduced by the microscopies, any differences in the macro results between the PLM and SHG analyses may also stem from how the macro deals with these differences. A macro that is influenced by factors in the images that are not linked with the characteristics to be analysed can easily produce false results. Also, its usefulness is limited, if in addition to the requirement of depicting the same kind of tissue, the images have to conform to further requirements in order to be analysed.

Ideally, therefore, the results of the PLM and SHG analysis should coincide, indicating that neither imaging differences, nor the way the macro deals with them, interfere with the correct identification and quantification of the underlying biological data. This demonstrates the degree of robustness of the macro and of the choice of image properties used for analysis.

A consequence of the different principles of image acquisition could also be that certain biological features that cannot be seen in one microscopy are evident in the other. In this case, the differences in the images contain biologically relevant information of the kind potentially important for tissue classification. Any difference in the analysis of the images could therefore reveal new information that is accessible by some imaging modalities only. Whether this is the case has to be assessed by closely inspecting the results on a case-by-case basis.

Fig. 5.8 illustrates how the microscopy images relate to the biological data. For explanation see figure legend. In reality, the entirety of the biological reality is only partially known, as is the subset of characteristics relevant for classification. The areas captured by one imaging modality but not by another are easier to determine by comparison. The chal-

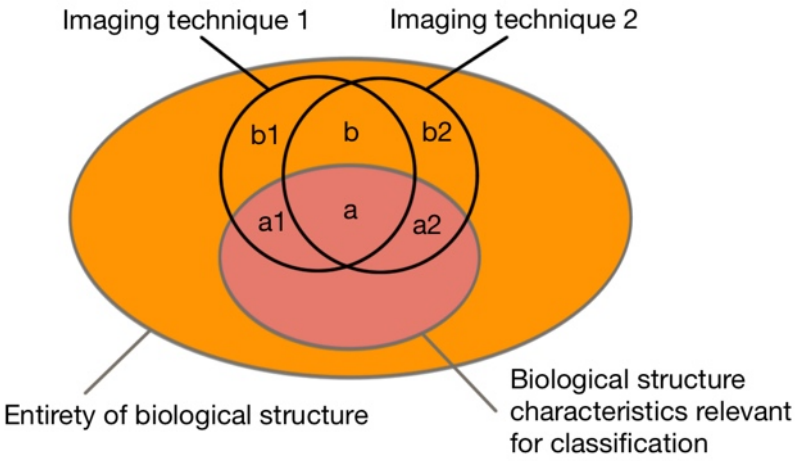


Figure 5.8: Schematic drawing of the theoretical relationship between different imaging modalities and the underlying structure. The basis of all data are the real biological structures present in living things. Only some of these structures will be markers for tissues, conditions, age stages or any other groupings. Any imaging modality will only be able to capture some aspects of the biological reality. Choosing the right modality will ensure that some of the data captured contain information on structures relevant for classification into groups, and therefore of interest for analysis (areas a, a1 and a2). Most likely, the imaging modality will also capture information that is not relevant for classification and analysis (areas b, b1 and b2). Each imaging modality, by nature of its different physical principles and tissue preparation requirements, will capture some of the same and some additional structures of the underlying biology. A robust and versatile analysis strategy will be most successful when the data to be analysed fall into area a, where the structure both yields information relevant for classification, and can be captured by more than one imaging modality. (Artifacts are not taken into account in this schematic, which would correspond to areas where the imaging modality circles reach into areas outside the biological structure.)

lenge is often when deciding on an analysis whether the quantity to be analysed belongs into the a or the b areas.

5.3 Results Comparison

When plotting the results of the same analysis as was done for PLM for the SHG images, the SHG graphs look very similar to those from the PLM analysis.

The results of the SHG analysis are shown in figs 5.9, 5.10 and 5.11. Following these, the results of the PLM analysis is shown again, for ease of referencing (figs. 5.12, 5.13 and 5.14, duplicate of figs. 4.27, 4.28 and 4.29).

As before, *AxanisAve* and *AngleR* in combination are sufficient to distinguish between groups A, B and C. *AxanisDev*, on the contrary, does not show sufficient differences between the groups to help in their identification. (While this is true for the tissue examined here, it can be envisaged that in different tissues the *AxanisDev* might show significant differences between the groups of classification.)

In the *AxanisAve* and *AngleR* graphs (figs. 5.9 and 5.11) the curves that allow differ-

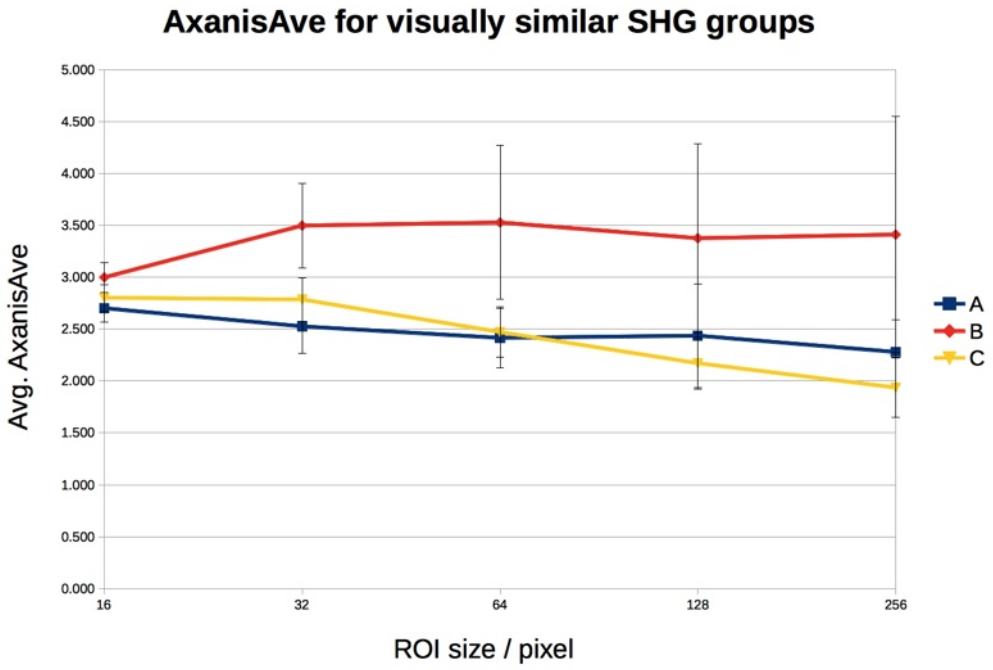


Figure 5.9: *AxanisAve* for SHG groups A, B and C. A: n = 3, B: n = 10, C: n = 9.

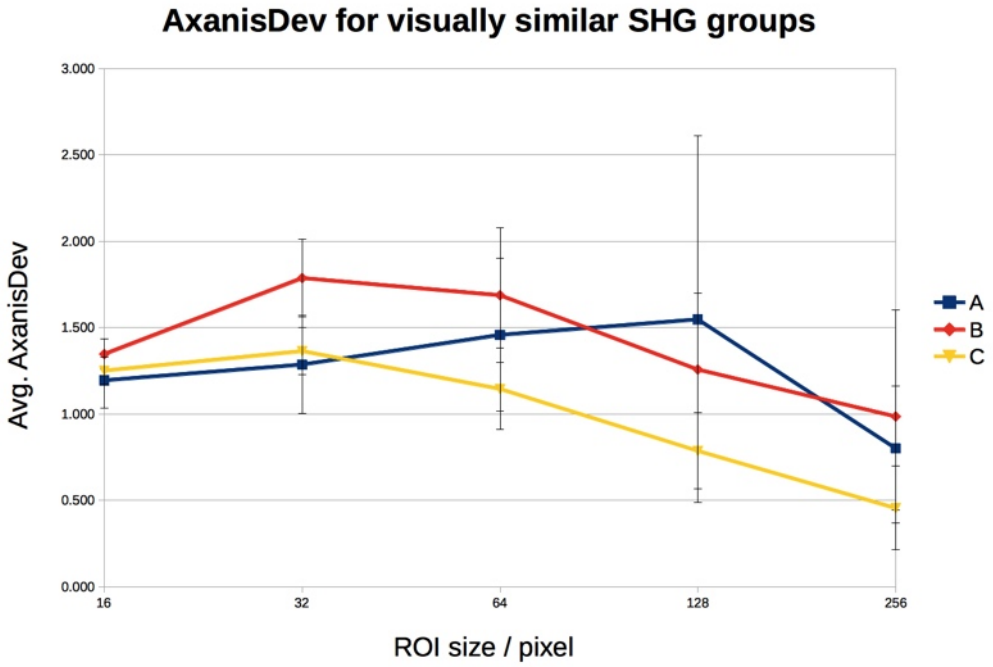


Figure 5.10: *AxanisDev* for SHG groups A, B and C. A: n = 3, B: n = 10, C: n = 9.

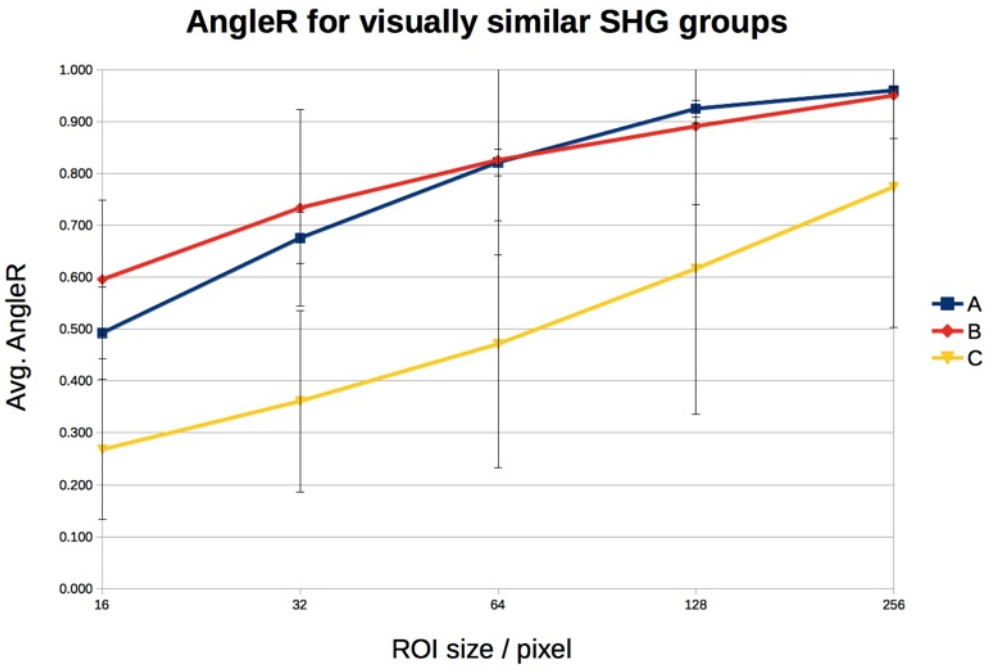


Figure 5.11: *AngleR* for SHG groups A, B and C. A: n = 3, B: n = 10, C: n = 9.

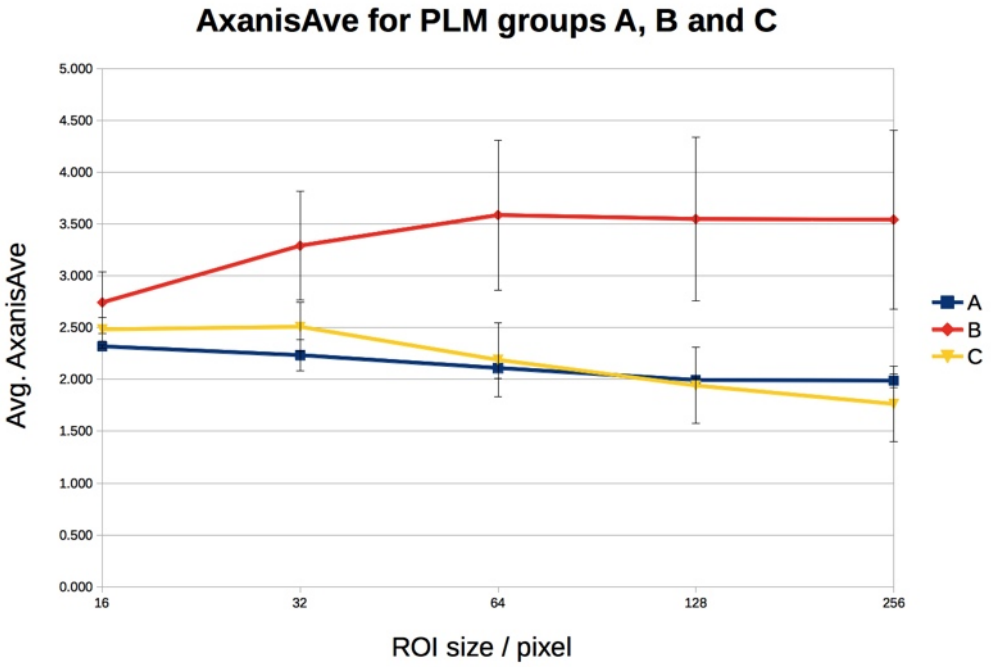


Figure 5.12: *AxanisAve* for PLM groups A, B and C. A: n = 3, B: n = 12, C: n = 8.

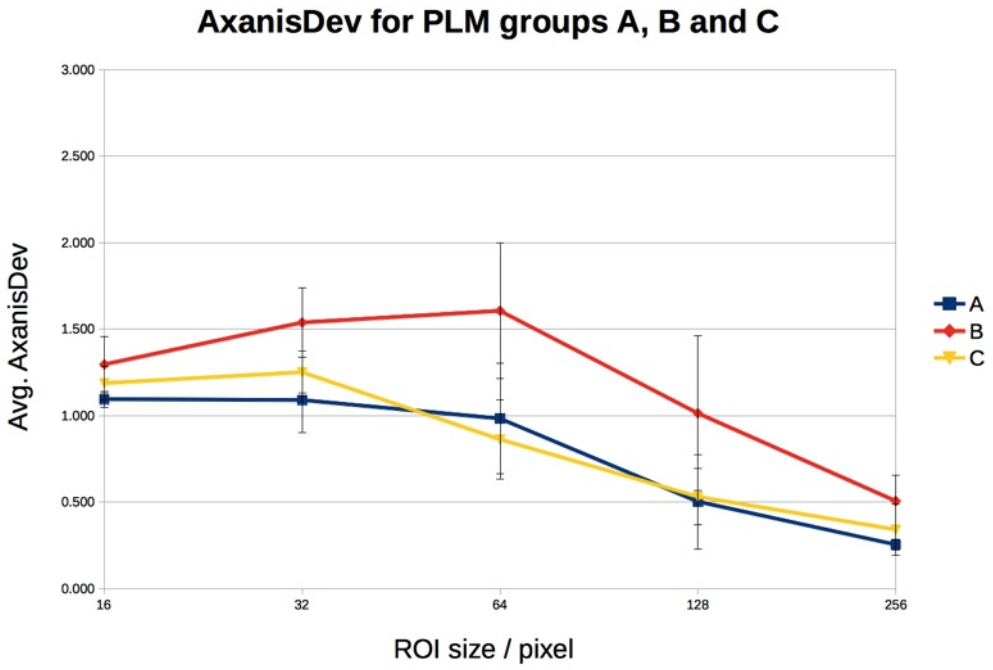


Figure 5.13: *AxanisDev* for PLM groups A, B and C. A: n = 3, B: n = 12, C: n = 8.

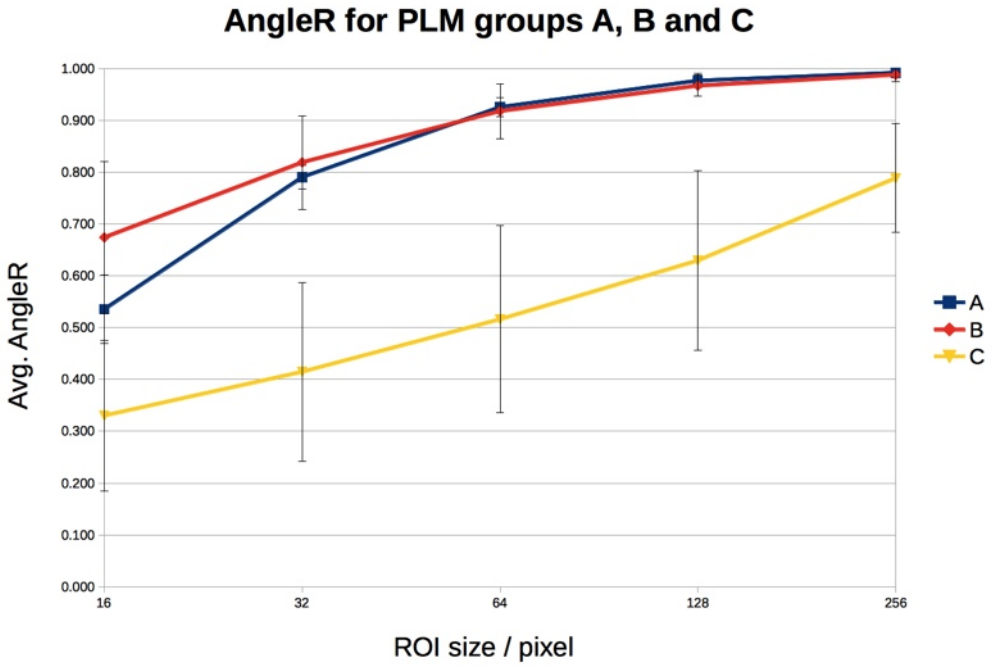


Figure 5.14: *AngleR* for PLM groups A, B and C. A: n = 3, B: n = 12, C: n = 8.

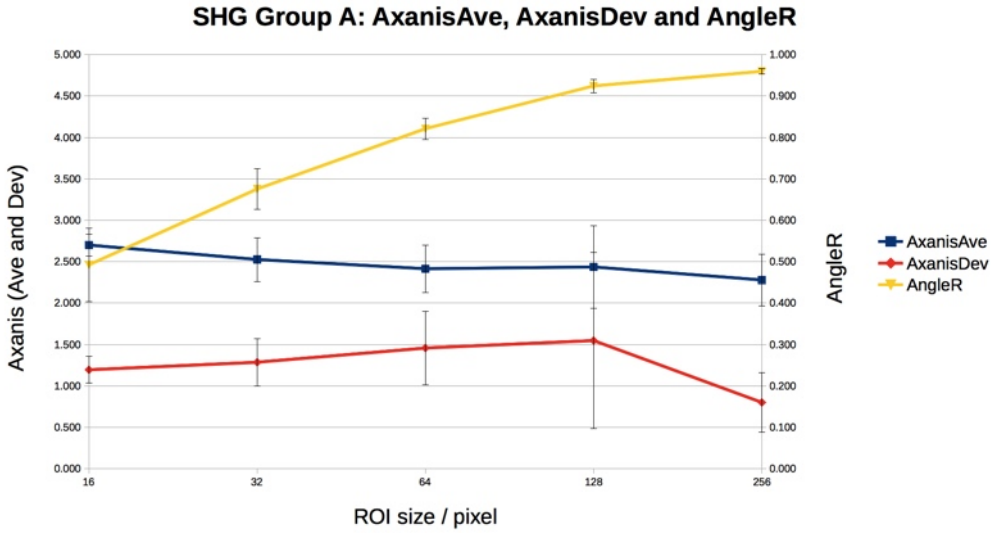


Figure 5.15: *AxanisAve*, *AxanisDev* and *AngleR* curves for SHG group A. Note that *AxanisAve* stays constant throughout all ROI sizes, *AxanisDev* decreases only slightly, and *AngleR* rises from a medium value to almost 1 at large ROIs. The error bars are standard deviations from the variation between images, n = 3.

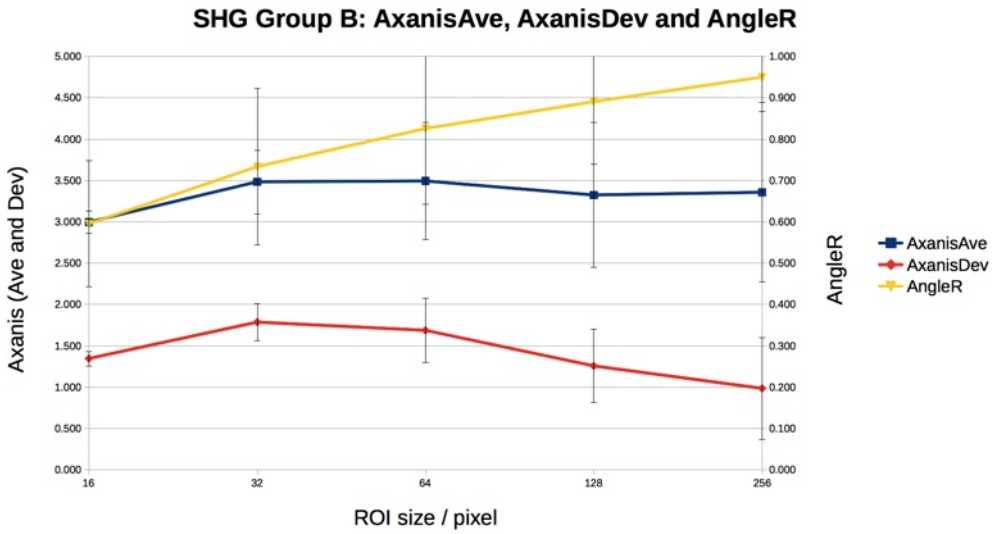


Figure 5.16: *AxanisAve*, *AxanisDev* and *AngleR* curves for SHG group B. Here *AxanisAve* is at first similar to A, but slightly rises at larger ROI sizes, *AxanisDev* is similar to A, and *AngleR* starts at a higher value than in A, but likewise rises to almost 1 at large ROIs. The error bars are standard deviations from the variation between images, n = 10.

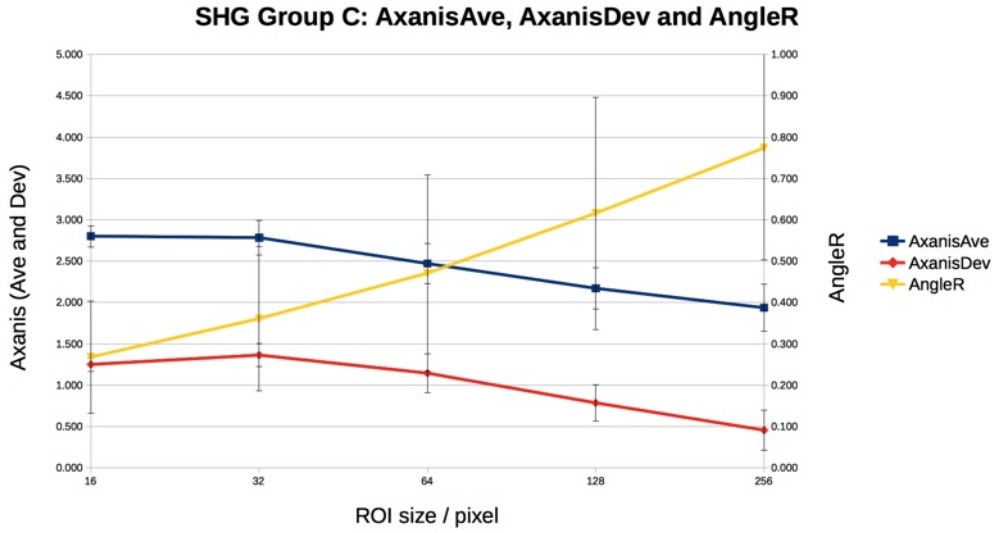


Figure 5.17: *AxanisAve*, *AxanisDev* and *AngleR* curves for SHG group C. The *AxanisAve* behaves very similar to A, and the *AxanisDev* curve follows the *AxanisAve* trend almost exactly. The *AngleR* begins at a lower value than in A and B and rises linearly to a value of only 0.8 at large ROIs. The error bars are standard deviations from the variation between images, n = 9.

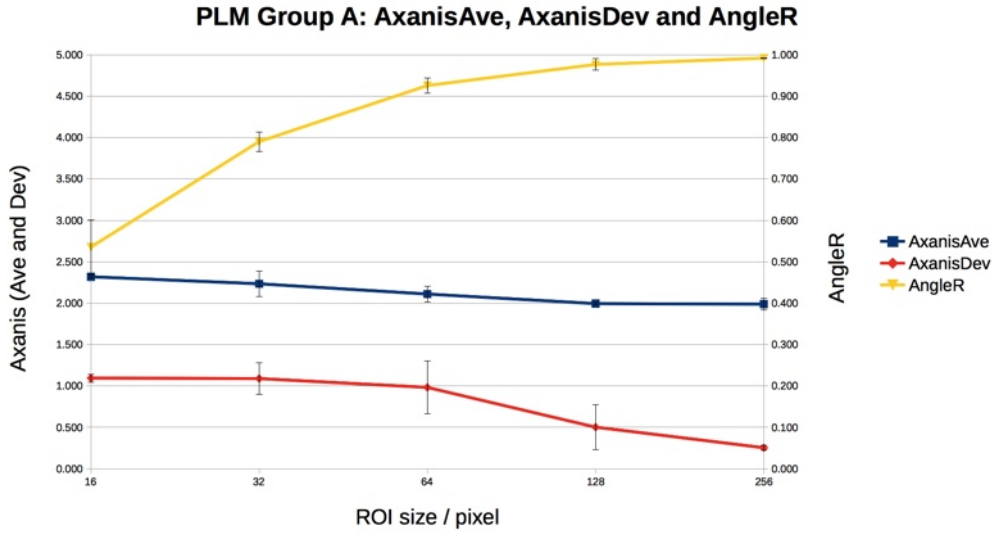


Figure 5.18: *AxanisAve*, *AxanisDev* and *AngleR* curves for PLM group A. Note that *AxanisAve* stays constant throughout all ROI sizes, *AxanisDev* decreases only slightly, and *AngleR* rises from a medium value to almost 1 at large ROIs. The error bars are standard deviations from the variation between images, n = 3.

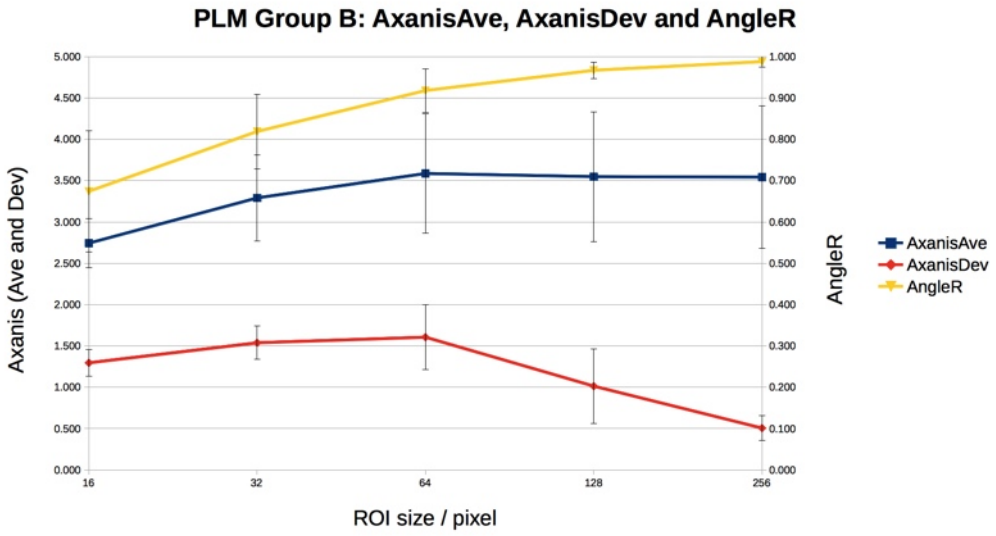


Figure 5.19: *AxanisAve*, *AxanisDev* and *AngleR* curves for PLM group B. Here *AxanisAve* is at first similar to A, but slightly rises at larger ROI sizes, *AxanisDev* is similar to A, and *AngleR* starts at a higher value than in A, but likewise rises to almost 1 at large ROIs. The error bars are standard deviations from the variation between images, n = 12.

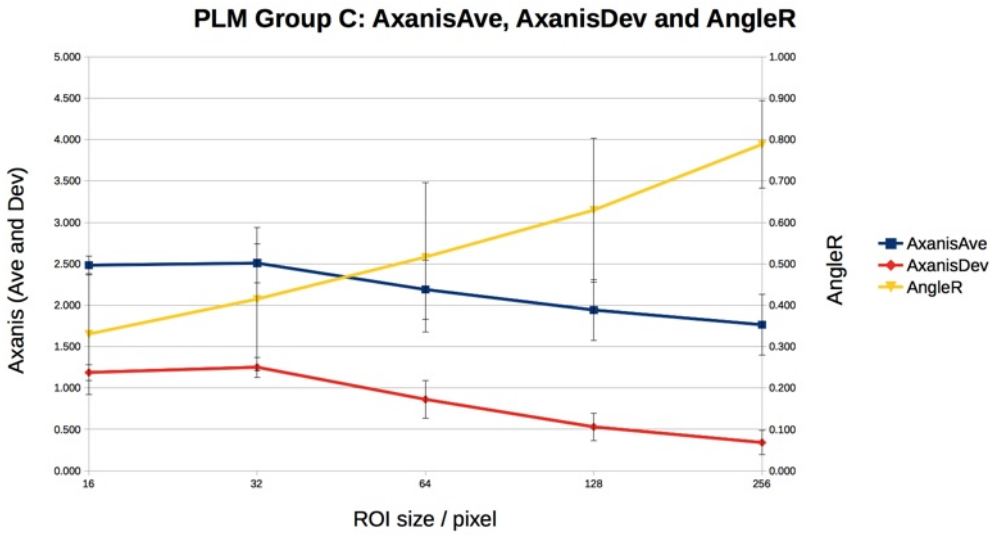


Figure 5.20: *AxanisAve*, *AxanisDev* and *AngleR* curves for PLM group C. The *AxanisAve* behaves very similar to A, and the *AxanisDev* curve follows the *AxanisAve* trend almost exactly. The *AngleR* begins at a lower value than in A and B and rises linearly to a value of only 0.8 at large ROIs. The error bars are standard deviations from the variation between images, n = 8.

entiation (B versus A and C for *AxanisAve*, C versus A and B in *AngleR*) are distinct at almost all ROI sizes. Therefore, ROI size is not critical in this analysis for distinguishing groups A, B and C. The scatter in the data points is generally larger for SHG images than for PLM images (compare for example fig. 5.11 with fig. 5.14). The finer structure visible in the thicker collagen fibres in the SHG images might account for this effect.

Since the shapes of the graphs are almost identical between PLM and SHG for *AxanisAve* and *AngleR* (and less so for *AxanisDev*) in spite of the scatter, it is likely that the shape of these graphs is characteristic for the image groups and not accidental. The fact that both *AxanisAve* and *AngleR* are necessary to be able to distinguish between groups A, B and C points to the nature of the problem not being trivial, as complex differences are hard to quantify in a straightforward and intuitively intelligible way. However, the concepts of *AxanisAve*, *AxanisDev* and *AngleR* are not unduly complex, while still capturing whole-image textural qualities that cannot be quantified by eye.

Also when comparing groups A, B and C in SHG and PLM, there are great similarities in *AxanisAve*, *AxanisDev* and *AngleR*. In both microscopies, the graphs describing *AxanisAve*, *AxanisDev* and *AngleR* within one groups show the same tendencies and even similar absolute values almost at every ROI size. The greatest differences are at small ROI sizes, consistent with the observation that SHG can resolve smaller fibre structure even within larger fibres. Group A is characterised by high values for *AngleR* and lower values for *AxanisAve* and *AxanisDev* in SHG (5.15) and PLM (5.18), group B shows high values for *AngleR* and also for *AxanisAve*, but lower values for *AxanisDev* in SHG (5.16) and PLM (5.19), whereas group C demonstrates low values for *AngleR* and lower values for *AxanisAve* and *AxanisDev* in SHG (5.17) and PLM (5.20).

Hence, the analysis results do not show significant changes between the original microscopy technique for which the algorithm was developed (PLM) and another microscopy technique (SHG) producing qualitatively similar but not identical image data. The algorithm is therefore not only insensitive to the differences in microscopy technique, but also yields the same patterns associated with the groups.

5.4 Future work

The analysis of the SHG and PLM images using the strategy presented here yields very similar results. Thus it has been shown that the macro is not sensitive to the differences inherent in the imaging modalities and is capable of analysing either images. Having demonstrated robustness at this level, it would be interesting to use more diverse images to test the capabilities of the analysis strategy and macro further. It would be possible to use the analysis on images taken at different magnifications and compare the results, both between the PLM and SHG microscopies at higher magnifications and between the magnification levels.

SHG is capable of visualising tissue structures in 3D, producing stacks of scanned 2D images. Using the analysis presented here on these image stacks would make it possible to build up a “line spectrum” for each of the metrics, and ROI sizes, with each position on the “spectrum” originating from one of the stack images. The combination of such

spectra for all metrics and ROI sizes might together yield similar classification criteria as the metrics graphs plotted against ROI size for the images shown here.

Additional images might be derived from sources outside microscopy, as long as the properties (intermeshed fibres of varying thicknesses) that are the target of the analysis method are present.

In order to further characterise the macro it might be interesting to create simulation images with known characteristics of fibre length, distribution of thicknesses, directionalities and intermeshing, at different length scales, and subject them to the analysis. This might provide more direct information on the response of the *AxanisAve*, *AxanisDev* and *AngleR* to certain defined image characteristics, and of their individual sensitivities to those characteristics.

The wider purpose of the analysis is its usefulness in classifying large numbers of images by means that are not obvious by visual inspection. In this respect, the future direction of the work that was already discussed in the last chapter: transferring the algorithm into Pascal, finding a suitable method of adapting the threshold to the particle in the ROI and applying the analysis to a large population of images, still applies just as much to this chapter.

5.5 Conclusion

The same algorithm that was successful in the analysis of PLM images is shown to be equally successful in analysing images obtained by SHG microscopy. The patterns in the three metrics characteristic for the visually identified groups in PLM are replicated closely when the analysis is applied to SHG images and show the same structures for each group. This confirms the wider applicability of the algorithm due to its robustness to a different microscopy technique, and points to the possibility for further development of the algorithm for other complex types of image data. Automatic analyses like the one presented here make it feasible to deal with the increasing volume of image data produced by modern microscopy techniques.

Chapter 6

Transmission Electron Microscopy

In this chapter an example of a problematic analysis is presented that was ultimately unsuccessful. Attempts to characterise the image noise and quality are described and a method is suggested that allows a quantitative evaluation of the goodness of the analysis method or the image quality through exploration of the available parameter space.

6.1 Principles of Transmission Electron Microscopy

The quest to be able to see details on smaller and smaller scales has driven the development of biological tools from the invention of the first light microscope. However, the magnification of a good light microscope does not usually suffice to see the details necessary to understand much of biological function at the cellular level. Resolving power is limited theoretically by the wavelength of the light used, and at a wavelength of around 500nm the resolution limit, or smallest distance at which two points can still be discerned as separate, Δx, is about 150 nm (after Abbe's formula $\Delta x = \lambda/2A$, where λ = wavelength and A = numerical aperture of the lens).

In order to increase the resolving power further, in electron microscopy, accelerated electrons are used instead of light, through which a shorter wavelength can be achieved. The electrons are focussed magnetically into a beam, similar to the lens system in a light microscope, and then (in transmission electron microscopy) sent through the specimen that has been prepared in such a way that interesting features have been rendered electron-dense by staining with heavy metal salts. In this way, the magnifications that can be achieved reach a million times or more. The wavelength of the electrons is around 10^{-12} m when accelerated at 300 keV and the resolution limit is in the region of 0.2 Å. ([138]).

Although this resolving power allows imaging of much greater detail, the problem with biological specimens in particular is that they are by nature almost completely transparent to the electron beam, and that they are almost instantly destroyed by the beam energy and the high vacuum pressure within the microscope. Their structures can only be visualised by impregnation with heavy metal salts, and they require a number of substantial preprocessing steps that only indirectly indicate the different biological structures. This harsh treatment can often be a source of artifacts, and it is sometimes difficult to ascertain whether an image feature arises from a true biological structure. When the biological sample is not in optimal condition, already contains debris or its structure has been degraded by delays in fixation or long storage, the interpretation of the images becomes particularly

challenging.

Fig. 6.1 shows the principle of a Transmission Electron Microscope in a simplified schematic drawing.

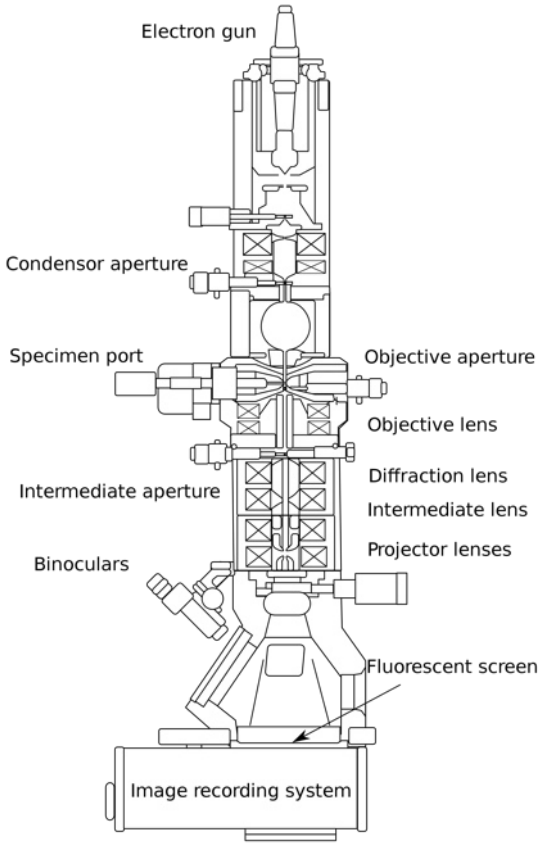


Figure 6.1: Schematic drawing of a Transmission Electron Microscope. An electron beam is accelerated and focussed onto a biological specimen stained with heavy metals. The unscattered, transmitted electrons are visualised by a fluorescent screen or CCD camera below. (Adapted from [139].)

6.2 The Images

TEM images of collagen in porcine sclera were kindly made available by Dr Riaz Akhtar from the Department of Mechanical Engineering at the University of Liverpool. Porcine sclera had been stained with uranyl acetate by following a standard protocol. The images were taken on a FEI Tecnai Spirit 120KV electron microscope by a member of Dr Akhtar's group.

We were given a total of 127 images, with magnifications varying from 16,500 to 87,000 times. The set of images contained collagen in cross and longitudinal sections, sometimes in separate images and sometimes in the same image. A number of the images also contained structures other than recognisable collagen.

Images that contained mainly non-collagenous structures were identified and excluded. Furthermore, we concentrated only on the images of collagen in cross-section, since cross-sectional and longitudinal images require a different analysis approach. For the develop-

ment of an analysis strategy, the 26 images of collagen in cross-section that met these criteria were chosen.

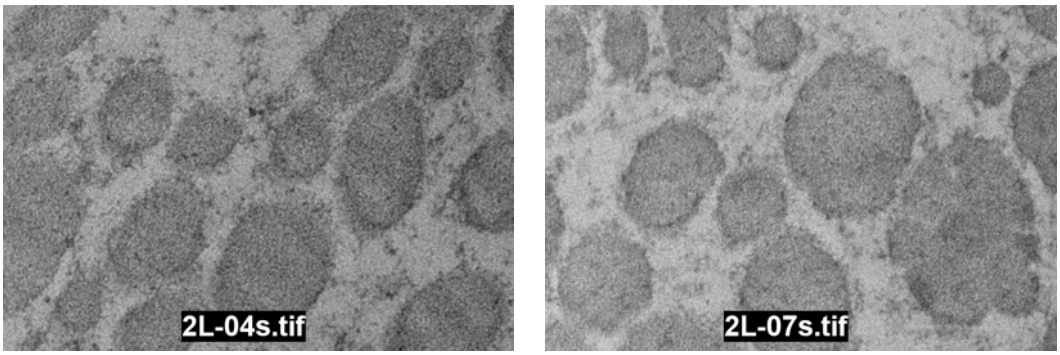
The image quality varied considerably, with occasional loss of the boundaries between collagen and the background (background is used here in the sense of the image areas that are not occupied by collagen structures) and substantial noise in the images. This problem could be due to a number of factors (Marion Pope, School of Veterinary Science, University of Liverpool; personal communication): tissue degradation before staining, degradation of staining solutions, incomplete washing during the staining procedure, or a combination of these.

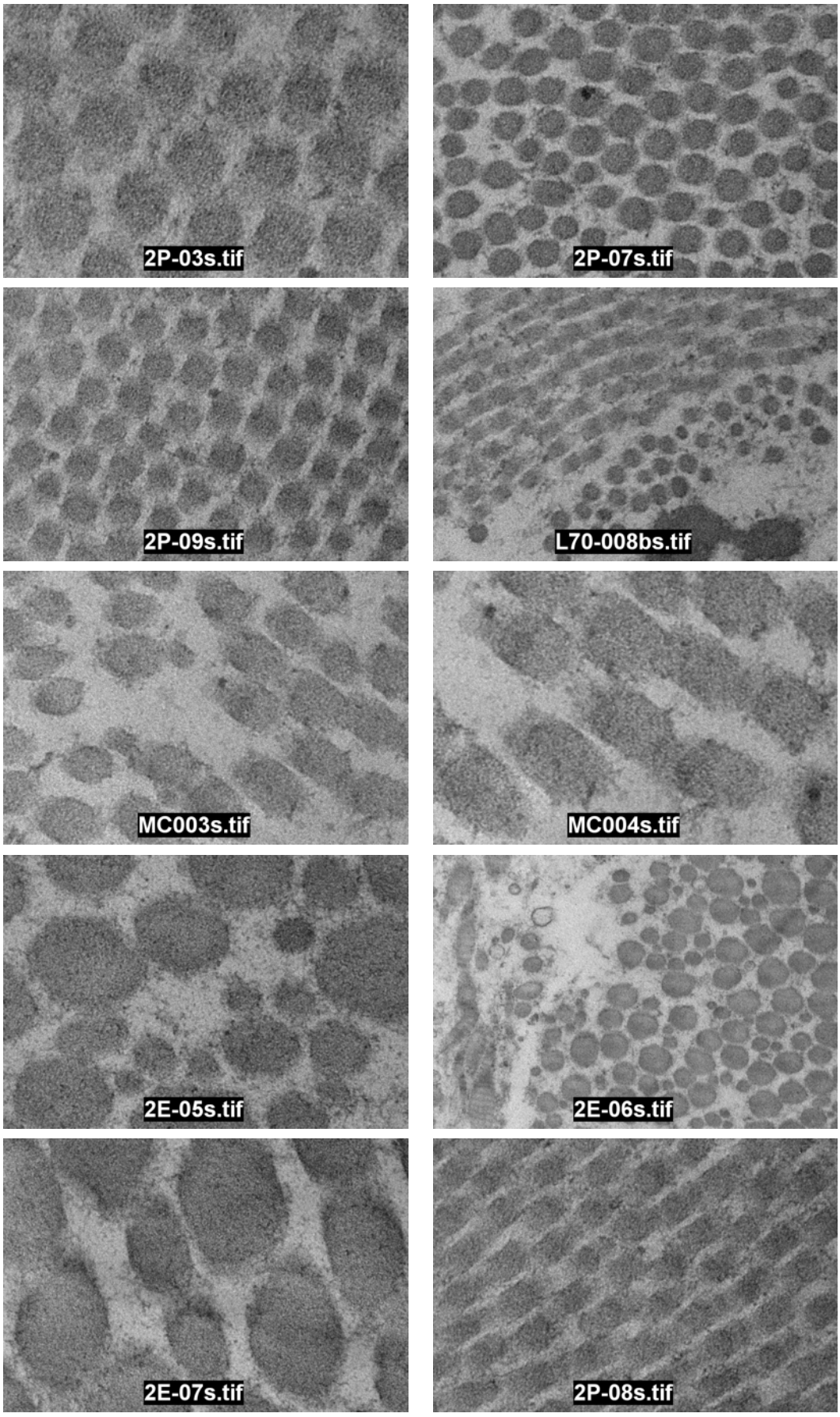
Since constraint on resources did not allow to acquire better quality images, analysis was attempted on this set of images. An added consideration in this was that it would be advantageous for an analysis to be able to deal with sub-optimal quality images, since in real life situations the ideal conditions for tissue preparation and processing available in a research setting may not always be feasible. Ideally, for TEM preparation, a biological tissue should be immersed in fixative within no more than a minute of death.

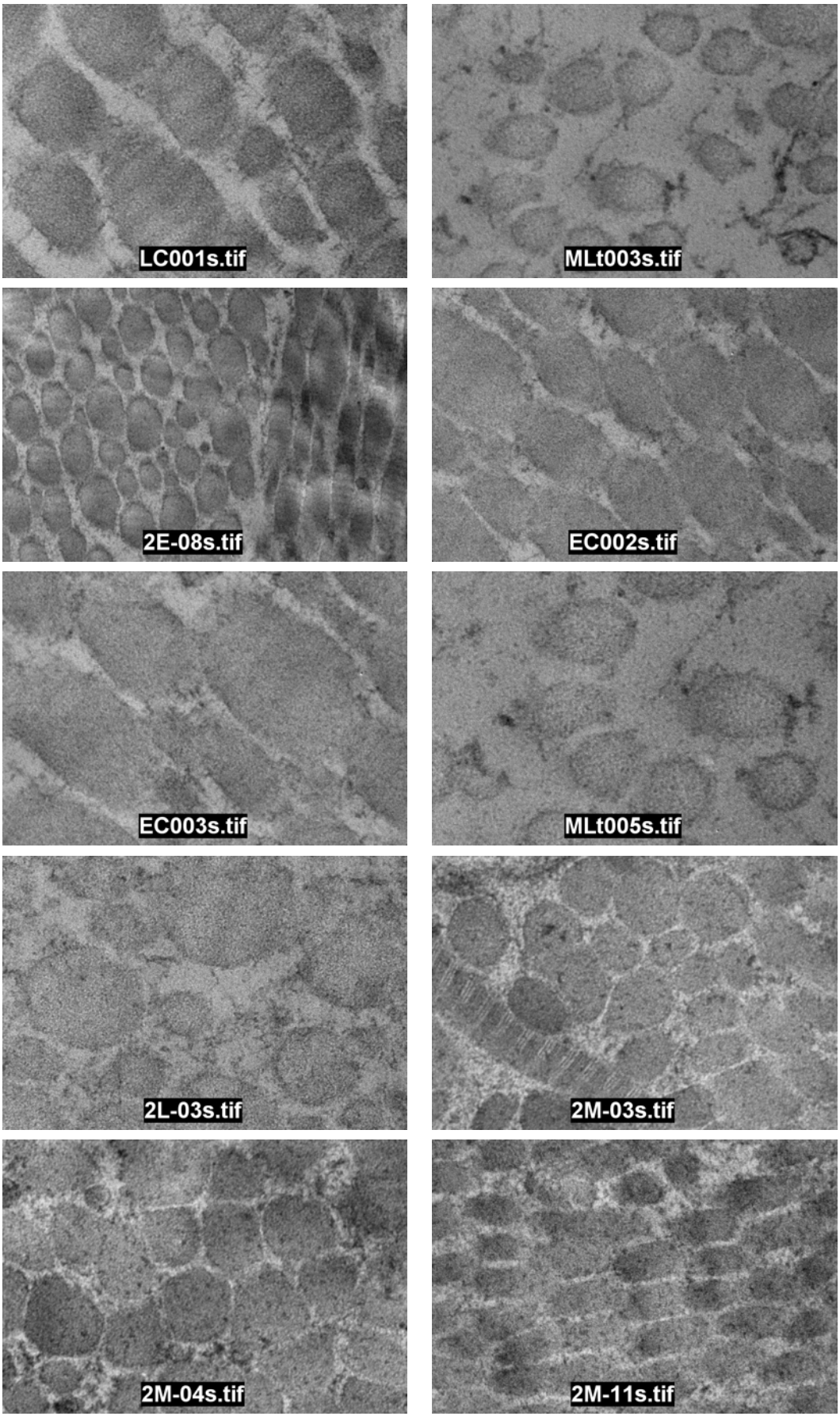
Whereas animal tissue such as porcine eyes is convenient to study, due to its similarity to human eyes and relatively easily availability, the study of human tissue is hampered by limited availability and variable life history. It is nevertheless necessary to investigate human tissue for confirmation of and improvement on animal studies. It may be possible to acquire animal tissue seconds after death in a research lab, but it is not usually possible to gain access to such fresh human tissue. Therefore, any TEM studies of analogous human tissue will likely be affected by a degree of tissue degradation in most cases. An image analysis strategy which still works on sub-optimal images would therefore be useful in practice, and the attempt to analyse these sub-optimal images has merit because of this fact.

A 871 x 589 pixel section was selected from each of the usable images (initial image size around 2400 x 1800 pixels) and duplicated into a new image. This allowed selection of a more uniform image, and shortened the processing time.

The image sections are shown in fig. 6.2. The images are divided into groups from *Green* through *Yellow* to *Orange* and *Red* according to their quality with regard to contrast and clarity of features. This classification arose after extensive experience with attempts at analysis.




2P-03s.tif
2P-07s.tif
2P-09s.tif
L70-008bs.tif
MC003s.tif
MC004s.tif
2E-05s.tif
2E-06s.tif
2E-07s.tif
2P-08s.tif


LC001s.tif
MLt003s.tif
2E-08s.tif
EC002s.tif
EC003s.tif
MLt005s.tif
2L-03s.tif
2M-03s.tif
2M-04s.tif
2M-11s.tif

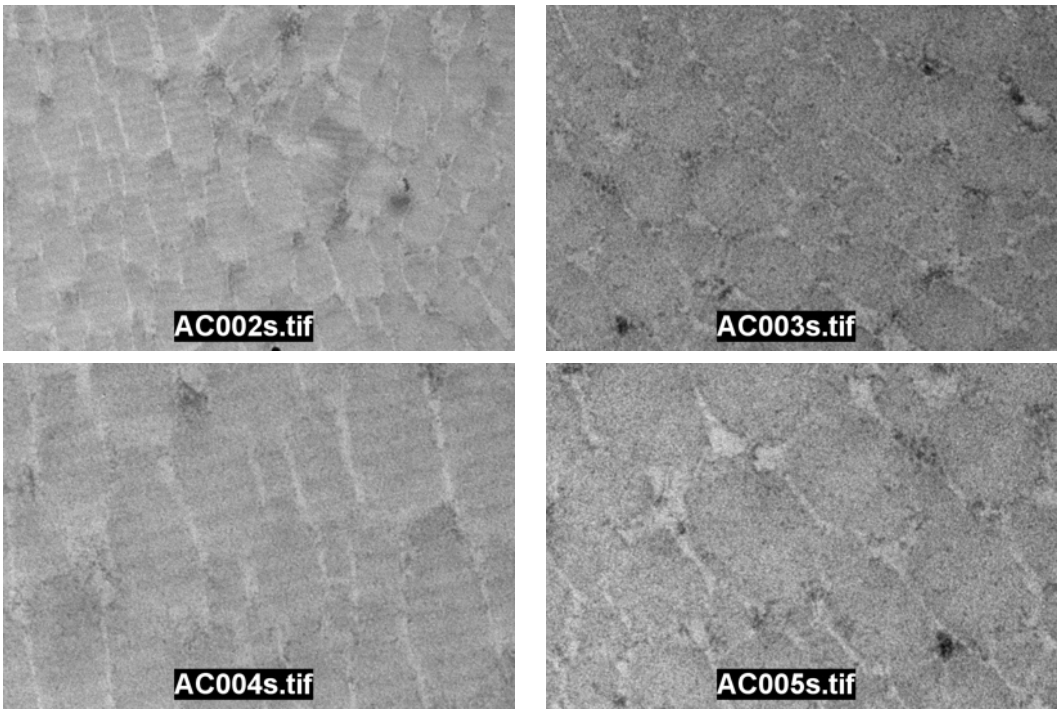


Figure 6.2: All TEM images for analysis of cross sections: Images were classed by quality: From 2L-04 to MC004 *Green* images (best), from 2E-05 to MLt003 *Yellow* images (somewhat worse), from 2E-08 to MLt005 *Orange* images (considerably worse), from 2L-03 to AC005 *Red* images (worst). The scales of these image differ, and due to the absence of calibration data in the file headers, the scales of the images could not be ascertained.

It can be seen that there is wide variation between the images in levels of contrast, degree of "graininess", sharpness of boundaries of the image features and angle at which the collagen fibres have been cut. Also, the nature of the blobs seems to differ between images. (Blob is a technical term used in image analysis and is an acronym of binary large object. It usually refers to a group of connected pixels in a binary image that represents a feature of interest that is the target of image segmentation. For simplicity, blob is also used here for an object that we would like to separate from the background, even if it is not yet binarised.)

We made extensive attempts to analyse this set of images, approaching the analysis from various conceptual angles, employing a wide variety of preprocessing steps, combining different techniques and using different starting images of the set for the development of the analysis. However, the best results we achieved were analyses that worked for the image they had been developed on, but fell down for the other images in the set. Mostly, our attempts yielded an unsatisfactory approximation to an analysis of acceptable quality. Since detailing these attempts would take up a large part of this thesis and are of limited instructive value individually, they are not presented here.

The AFM, PLM and SHG images show high enough contrast and low levels of noise and artifacts, and therefore the analysis strategy could focus directly on the extraction of biologically relevant data. Since this is not the case with the TEM images, we attempted to identify the image characteristics that prevented successful analysis.

At least in the case of these images, where the idealised blob takes the shape of a

roundish oval, the factors that interfere with blob identification and segmentation are noise, an unclear or non-existent boundary and low contrast in the image.

The distinction between noise proper and artifacts is a fuzzy one. Whereas artifacts are usually larger structures concentrated locally, noise tends to be on a smaller length scale and distributed more uniformly over the image, but both negatively affect the segmentation of the image into blobs and background. Noise can take a variety of forms. So-called salt and pepper noise consists of individual black and white pixels, but in other noise the pixels can also take on other grey values, covering the whole spectrum of the histogram from 0 to 255, or only a part of it. Not only which grey values the noisy pixels adopt is of interest, but also how they are distributed over the spectrum of grey values. There is no limitation in principle to the nature of the distributions, as this depends in each individual case on set of circumstances that constitute the origin of the noise. How the noise is distributed can have important consequences for the success of the analysis, though. If, for example, the noise populates only the lighter grey values, whereas the blob pixels occupy only the darker areas, a simple thresholding step might be able to separate them. If the noise and the blob grey values are very similarly distributed, this is not possible.

Related to this is the problem of low contrast. Low contrast means that the histogram of the whole image is very narrow, providing little opportunity to segment the image by the criterion of grey values. Furthermore, locally in the image there will be few occasions where a dark pixel can be found directly adjacent to a light one, providing a step that is sufficiently unique to be recognised as a property of a boundary.

Unclear boundaries affect both the interpretation of structures as blobs and the efficacy of those analysis methods that use prior information to search for particular features in the image. The Hough Transform (see Appendix A.8), for example, can be set to look specifically for image features like lines or circles. It does this by a kind of voting procedure which takes into account the strength of the grey values lying on the putative feature, and if the edge of a feature disappears in a part of this length the votes are weakened accordingly. If the existing edges are clear, it may be possible to segment the partial blob in a satisfactory way. However, the shapes of such partial blobs could easily skew the overall distribution, if characteristics like roundness, eccentricity or length of circumference are measured. There are ways of dealing with this to a certain extent, like for example the Convex Hull, in which the blob is placed inside a line that is then shrunk onto the shape of the blob like a shrink wrap, and the ultimate shape of this line is then taken as the blob shape. This method effectively deals with small gaps in an otherwise largely entire and convex shape, but provides little help when large parts are missing or for non-convex shapes.

In order to gain a more detailed understanding of this set of TEM images, we attempted to characterise the features obstructing our analysis attempts.

6.2.1 Histograms of Entire Images

The first and most straightforward step in understanding image quality is to establish a histogram of grey values. Shown here are two images from the green and one from the red

group. (The two images from the green group represent two different categories of blob appearance and thus both images are shown.) Fig. 6.3 and fig. 6.4 are from the green group, and fig. 6.5 is from the red group. The insets shows the histograms of the images as shown in the figures. (The histograms are non-continuous, which most probably is caused by a prior contrast enhancement or by digital shrinking and expansion of the images.)

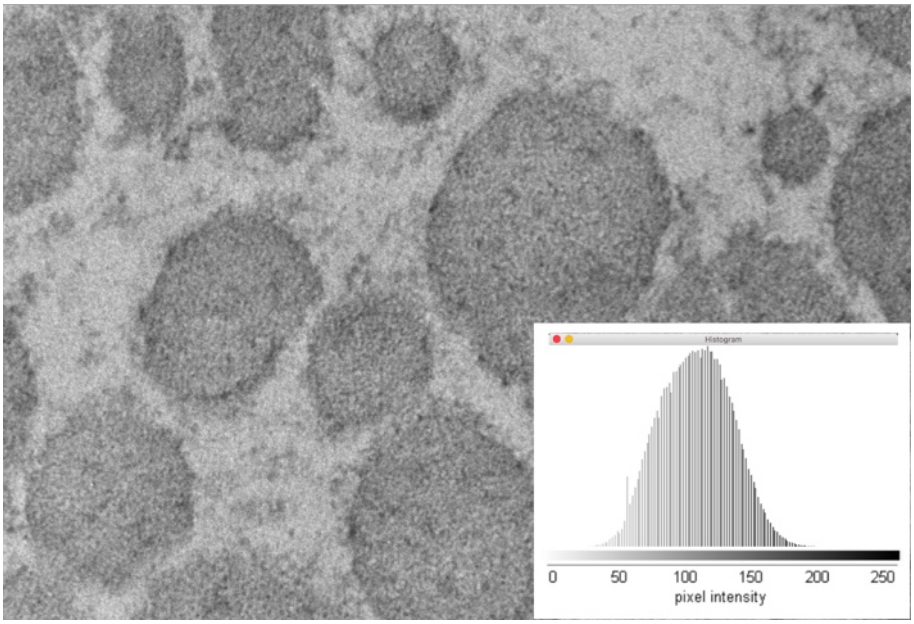


Figure 6.3: Image 2L-07 (green) and histogram of grey values. Note that the distribution of grey values is close to a Gaussian distribution with no additional peaks.

In neither of these representative images does the grey value distribution deviate noticeably from a Gaussian distribution. At most a small additional bulge can be found. This is the case also for the images not shown. This finding discourages the idea of successfully segmenting the image by direct thresholding.

6.2.2 Histograms of Segmented Images

Notwithstanding the shape of the whole-image histogram, it is still possible that two sharper peaks for background and blobs could be adding up to a vaguely Gaussian shape. In order to investigate this possibility, the same images as in the last section were divided into blobs and background by manually drawing a mask over the presumed blob shapes. Although identification by eye was not in all cases clear-cut, this still provides the best approximation of a good segmentation. In the following three figures, 6.6, 6.7 and 6.8, the original image is shown next to the manually drawn mask, followed by the image showing only the blobs and only the background, and underneath the histograms of the blobs alone and the background alone.

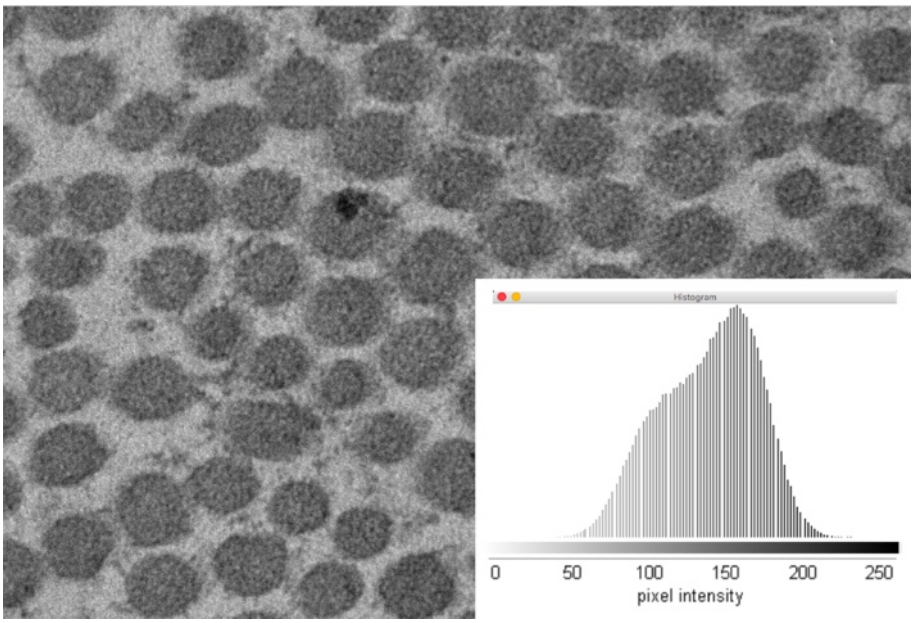


Figure 6.4: Image 2P-07 (green) and histogram of grey values. The distribution of grey values is still relatively close to a Gaussian distribution, but there is an additional expansion on the lighter slope.

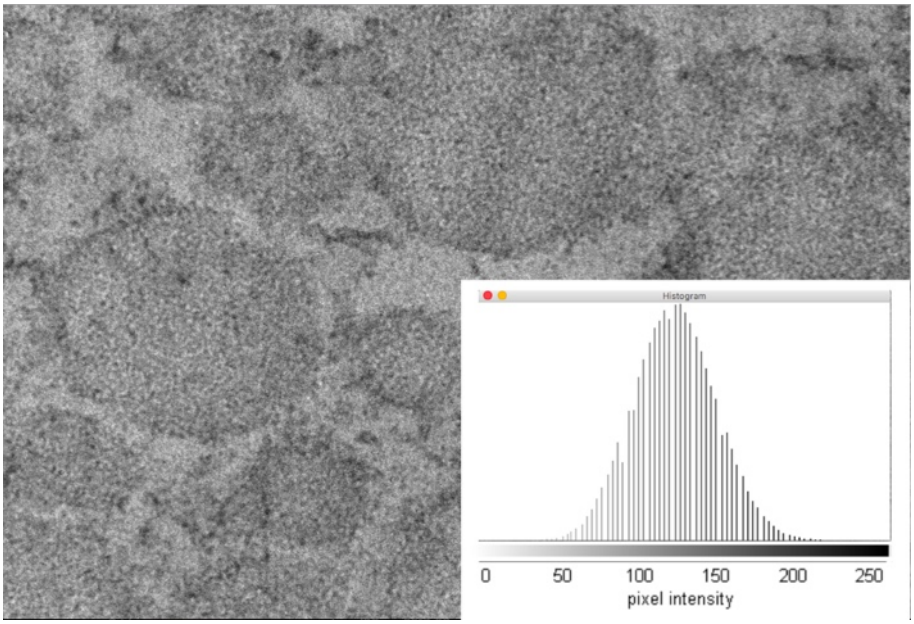


Figure 6.5: Image 2L-03 (red) and histogram of grey values. Also here the distribution of grey values is close to a Gaussian distribution with no additional peaks.

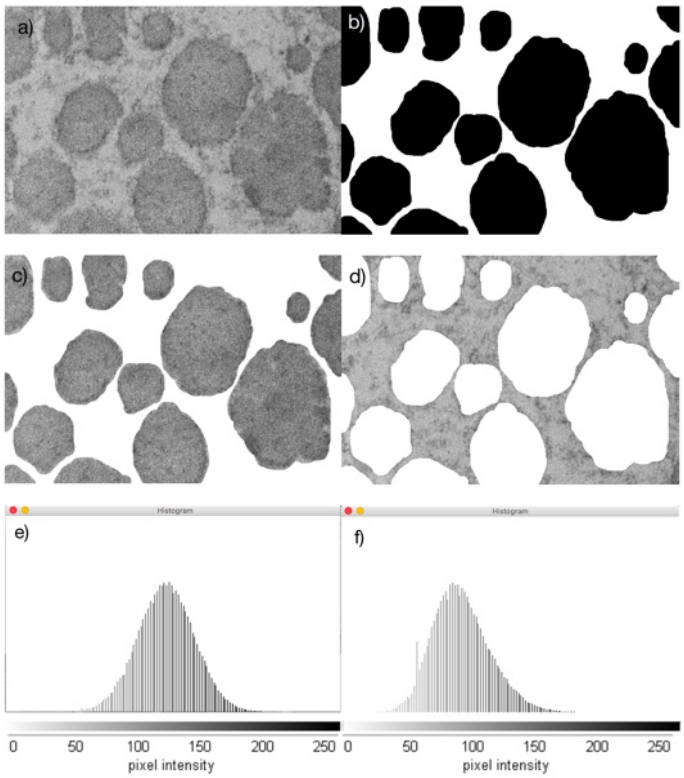


Figure 6.6: Image 2L-07: a) original image; b) manually drawn mask; c) image of blobs only with background removed; d) image of background only with blobs removed; e) histogram of blobs; f) histogram of background

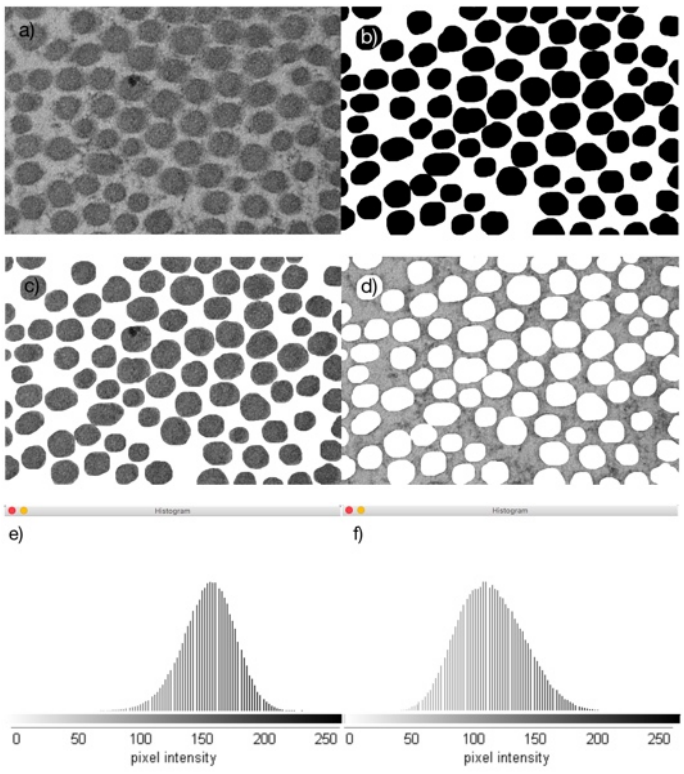


Figure 6.7: Image 2P-07: a) original image; b) manually drawn mask; c) image of blobs only with background removed; d) image of background only with blobs removed; e) histogram of blobs; f) histogram of background

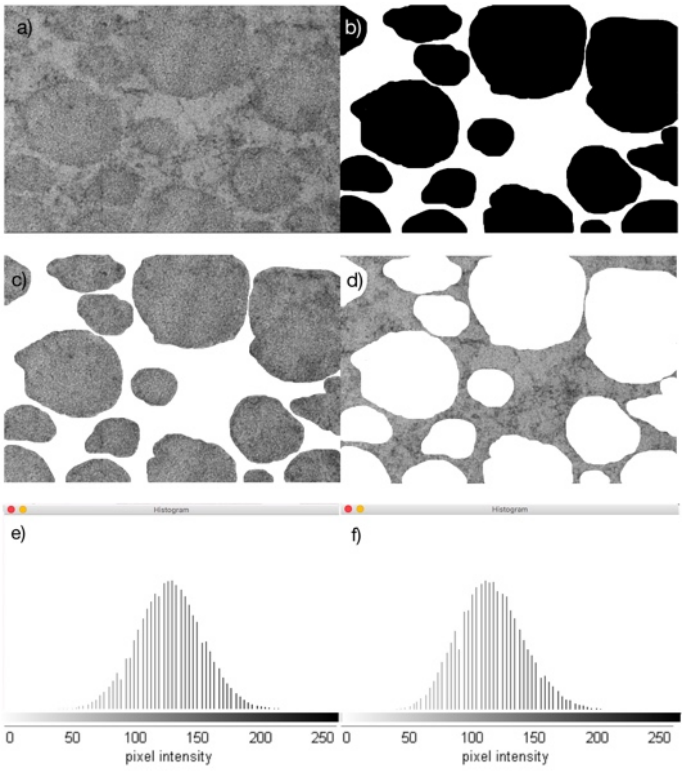


Figure 6.8: Image 2L-03: a) original image; b) manually drawn mask; c) image of blobs only with background removed; d) image of background only with blobs removed; e) histogram of blobs; f) histogram of background

Although this separation into blobs and background reveals that the two peaks do not exactly coincide in all three images, they are still very close, as was suggested by the whole-image histograms. This means that for the pixels in the centre of the peak whether they belong to a blob or to the background cannot be determined by grey value alone. Since there are a large number of pixels in this area, we decided to examine the boundary between blobs and background more directly.

6.2.3 Line ROIs Across Boundaries

The interface between a blob and the background is probably the most crucial element in the image that determines the success of any analysis. Even if the inside of the blob was almost identical to the background, as long as the outline was clear and distinguishable, a reasonable amount of accurate information about the blobs could be extracted in a relatively straightforward way. To get some more detailed information about this interface, two stretches of boundary were selected in each of the above images, one that was deemed to show a clear difference by eye, and one where the difference was not clear (fig. 6.9). A line ROI was drawn across the boundary and the corresponding plot profile was examined for clarity of the step that would show transition from background to blob. Since a thin

line ROI is more affected by noise and a thicker ROI averages over a wider number of edge pixels, the thickness of the ROIs was increased from 1 through 5, 10 15 to 20 pixels. The line ROIs were drawn at right angles to an imaginary tangent on the blob to avoid smearing out the edge through angle effects. Figs. 6.10, 6.11 and 6.12 show the "clear" boundary on the left and the "unclear" boundary on the right. The line ROI widths increase in descending order and each is shown with its plot profile next to it.

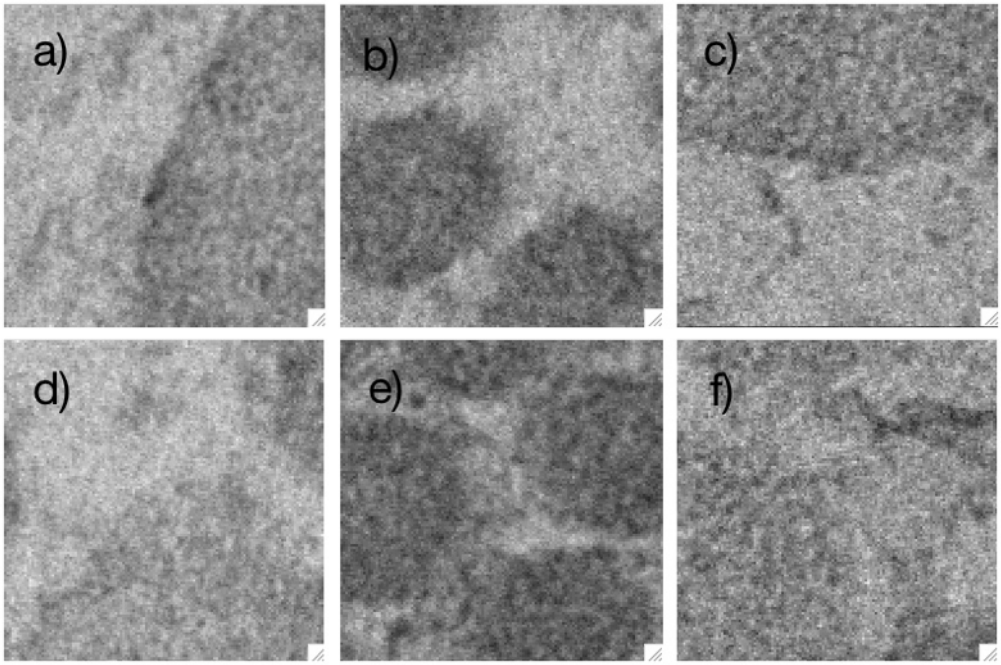
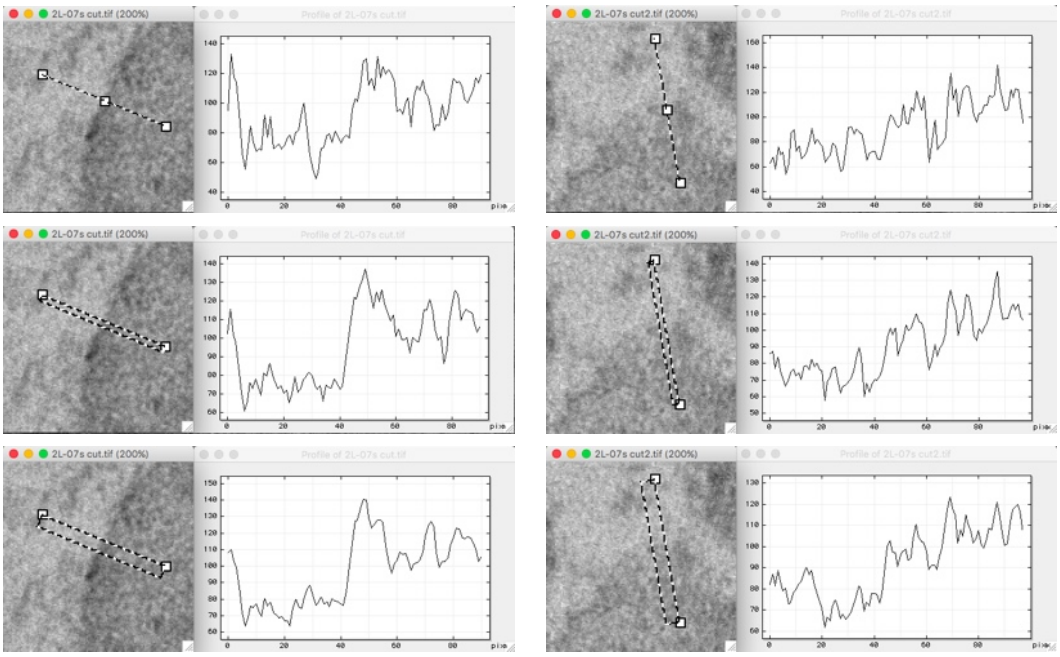


Figure 6.9: Boundaries between background and blobs: a), b) and c) clear boundaries; d), e) and f) unclear boundaries; a) and d) from image 2L-07; b) and e) from image 2P-07; c) and f) from image 2L-03



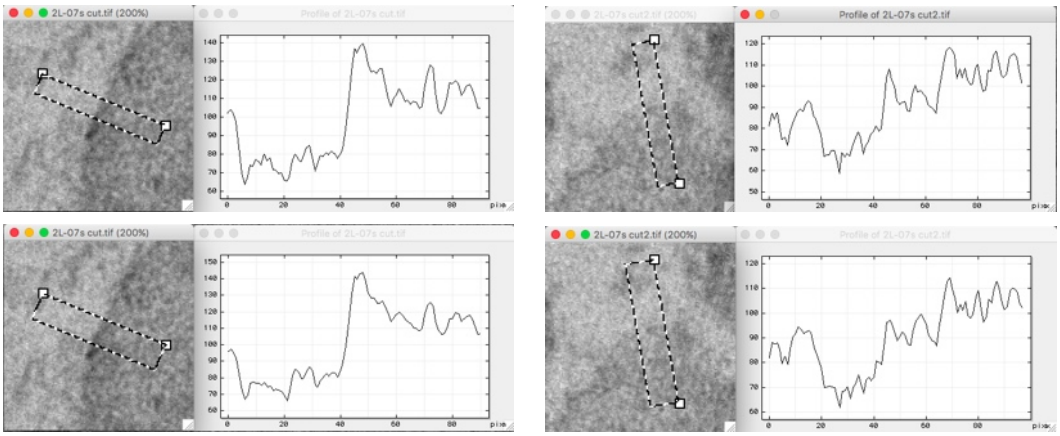


Figure 6.10: Image 2L-07: line ROIs of various thicknesses (1, 5, 10, 15 and 20 pixels) drawn through the background-blob boundary. First column shows the line ROI and plot profile through what is considered a clear boundary. Second column shows the same drawn through what is considered an unclear boundary.

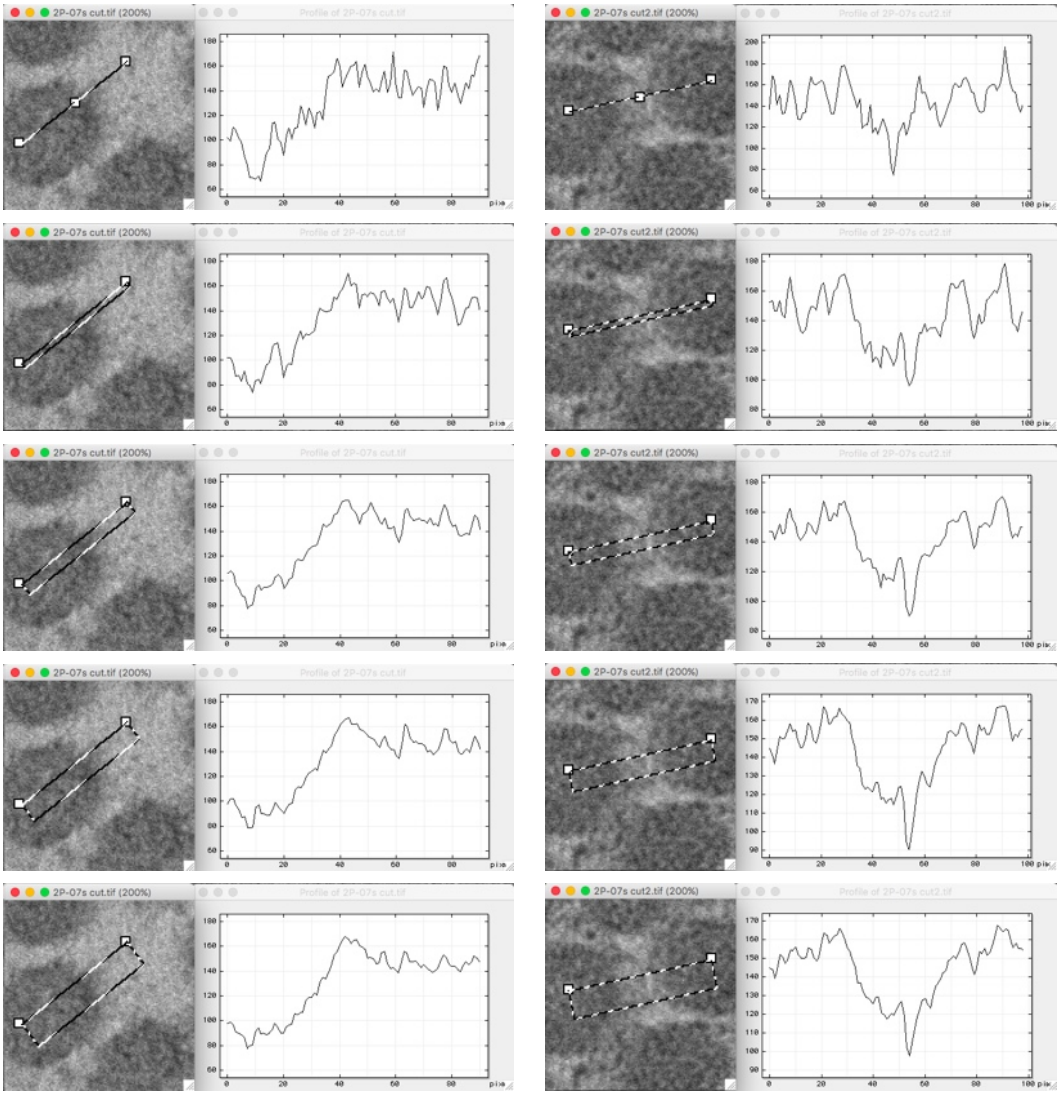


Figure 6.11: Image 2P-07: line ROIs of various thicknesses (1, 5, 10, 15 and 20 pixels) drawn through the background-blob boundary. First column shows the line ROI and plot profile through what is considered a clear boundary. Second column shows the same drawn through what is considered an unclear boundary.

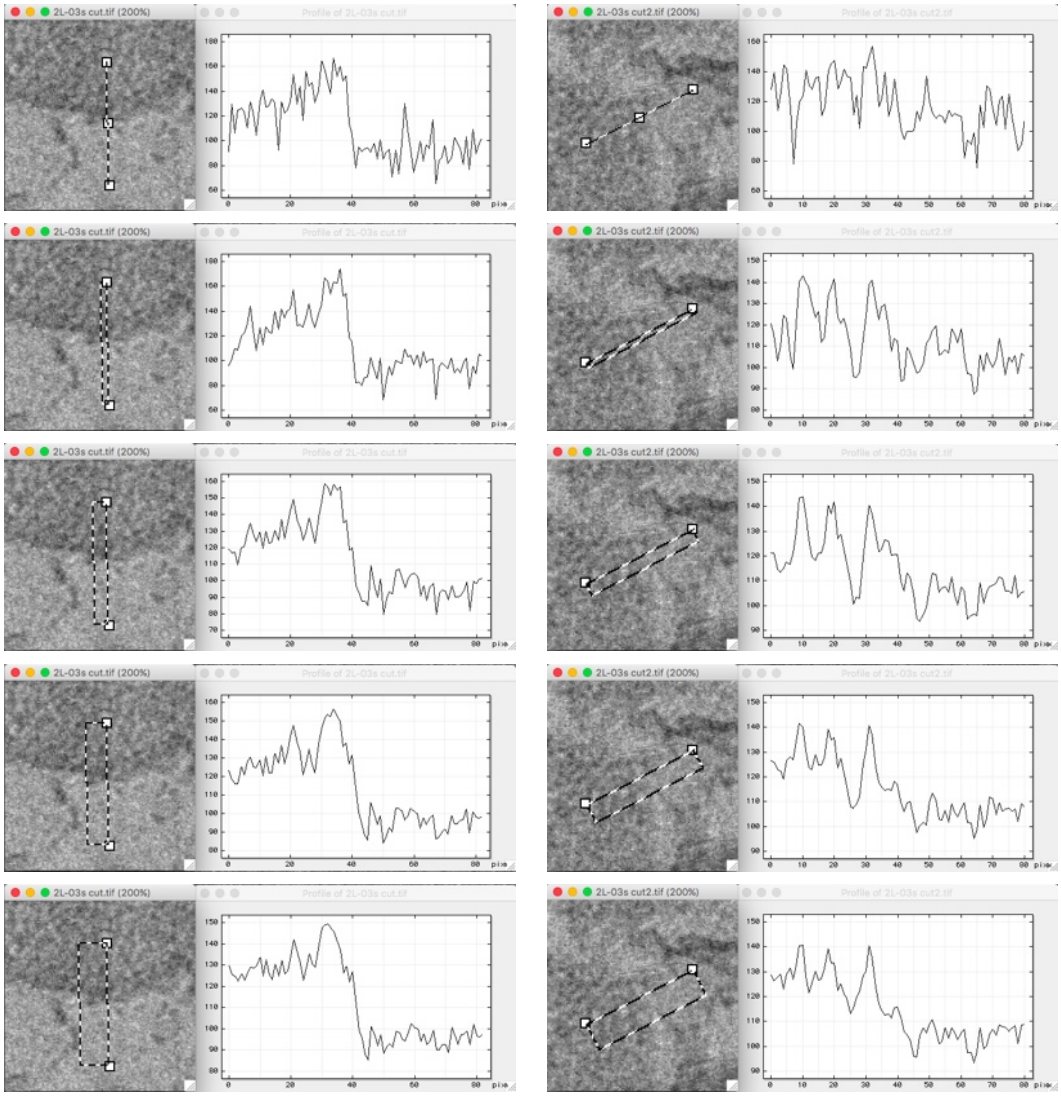


Figure 6.12: Image 2L-03: line ROIs of various thicknesses (1, 5, 10, 15 and 20 pixels) drawn through the background-blob boundary. First column shows the line ROI and plot profile through what is considered a clear boundary. Second column shows the same drawn through what is considered an unclear boundary.

Although noise is evident in all of these plot profiles, it is generally the case that wider ROIs generate plot profiles with more recognisable edge steps. As expected, the edges that seem clearer by eye also have plot profiles with clearer steps. Looking at the one-dimensional plot profiles alone it would in many cases not be easy to determine where blob and background meet, and the context of the two-dimensional image is necessary to gain more certainty. It would be difficult to find criteria that would allow a computer algorithm to reliably identify edges from this variety of plot profiles. The steep slope of the edges in fig. 6.10 and fig. 6.12 is very different, for example, from that in fig. 6.11, which is much shallower. Also, the height of the highest edge step is only about twice that of the noise

level in these examples, and such low contrast in an image makes it likely that there are many areas where the steps in the plot profiles disappear through random noise.

6.2.4 Fourier Filtering

One very powerful method for removing noise from images is through Fourier filtering. The FFT can deconstruct an image into a representation in frequency space, resulting in a phase and a power image. The phase image is not usually shown, as it does not contain information relevant to image manipulation. The power image is also often called the FFT image and contains information about the "power" of a certain frequency in the original image, that is "how much" of that frequency component there is. The FFT image shows the presence of image elements as a function of their scale (= "wavelength" in pixels, which translates into frequency). This operation can be reversed, and through an inverse FFT the original image is restored. Manipulation of the FFT image allows elements in the original image to be targeted through their distinct scale or frequency. For detailed examples see [114].

If noise is present only at a certain frequency, it can often be removed by deleting that frequency region from the FFT image. The inverse FFT image will be the original without any elements of the deleted frequency. When all frequencies higher than a threshold frequency are removed, this is called a low pass filter. When all frequencies lower than a threshold frequency are removed, it is called a high pass filter. (Note that in this context threshold refers to the cut-off point in frequency in the FFT filter, and not, as usual, to the binarisation of an image by rendering grey values above and below a certain threshold grey value into black and white.)

For the sake of convenience, the scale in pixels of image elements is referred to here as "frequency" as well, although the frequency is really image size in pixels / size of image element in pixels, that is 1 / number of pixels.

We applied Fourier filtering in an attempt to improve our image set. Fig. 6.13 shows a plot profile of the noise in blobs and in the background in one of our images. The smallest noise appears to be on the scale of around 5 pixels.

Although it is possible that noise at larger scales is also present, we first tried to remove the smallest scale noise by FFT filtering. Fig. 6.14 shows the original and its FFT image. No asymmetric or radial features are visible by eye in the FFT image, apart from two small spots at about the 11 and 5 o'clock positions, which originate from electrical noise in the microscope. This might indicate that there may not be one distinct noise scale in the original image.

Fig. 6.15 shows how the low pass and high pass filters are applied to the FFT image and the resulting inverse FFT images. (Image SXM allows the user to draw a circle into the FFT image, guided by the information of frequency or pixel number at the current mouse position in the info window, and then applies the appropriate operations to implement the filter.)

The images b) and d) in fig. 6.15 are separated by frequency components, and therefore, if the frequency elements above and below the threshold were characterised by distinct

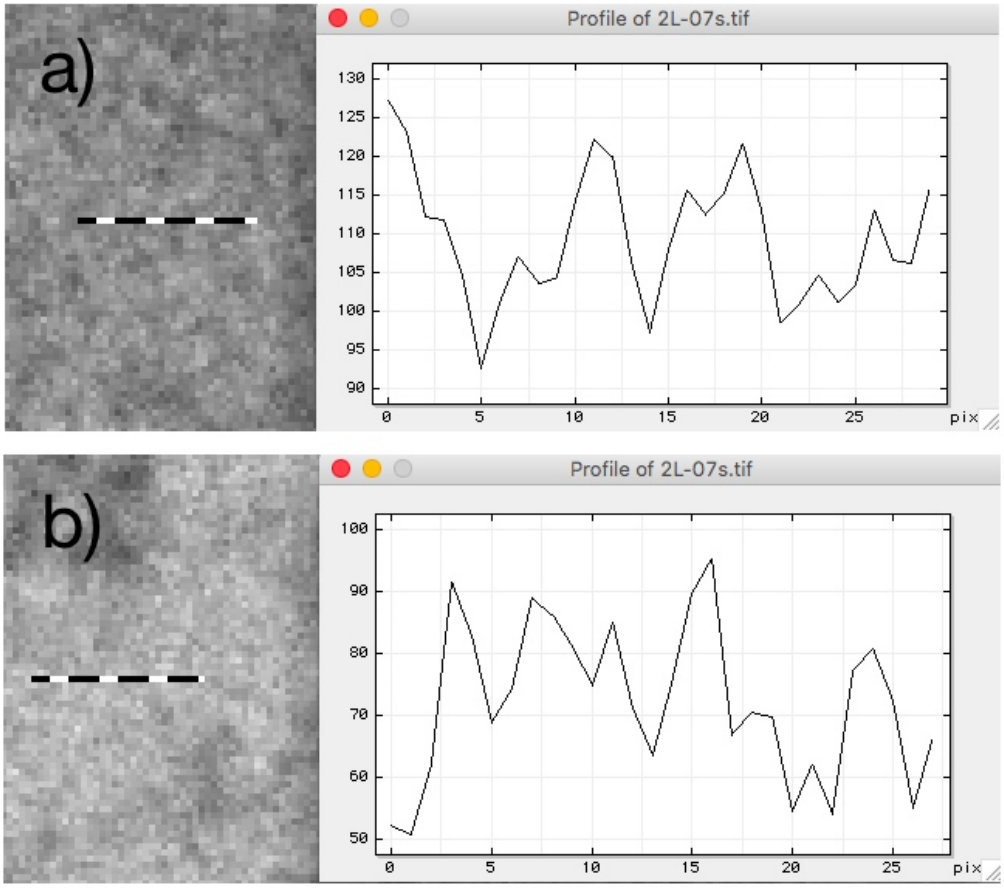


Figure 6.13: Determination of the length scale of the noise. a) ROI drawn inside blob with plot profile. b) ROI drawn in background with plot profile. The smallest noise in on the length scale of approximately 5 pixels.

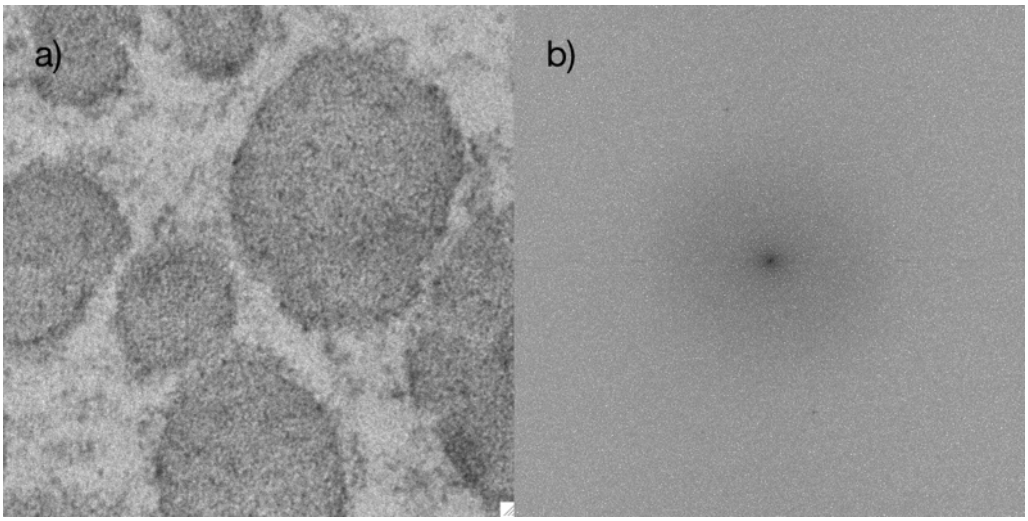


Figure 6.14: An original image and its FFT image. There are no dark areas in the FFT image that might indicate a particularly prevalent noise frequency.

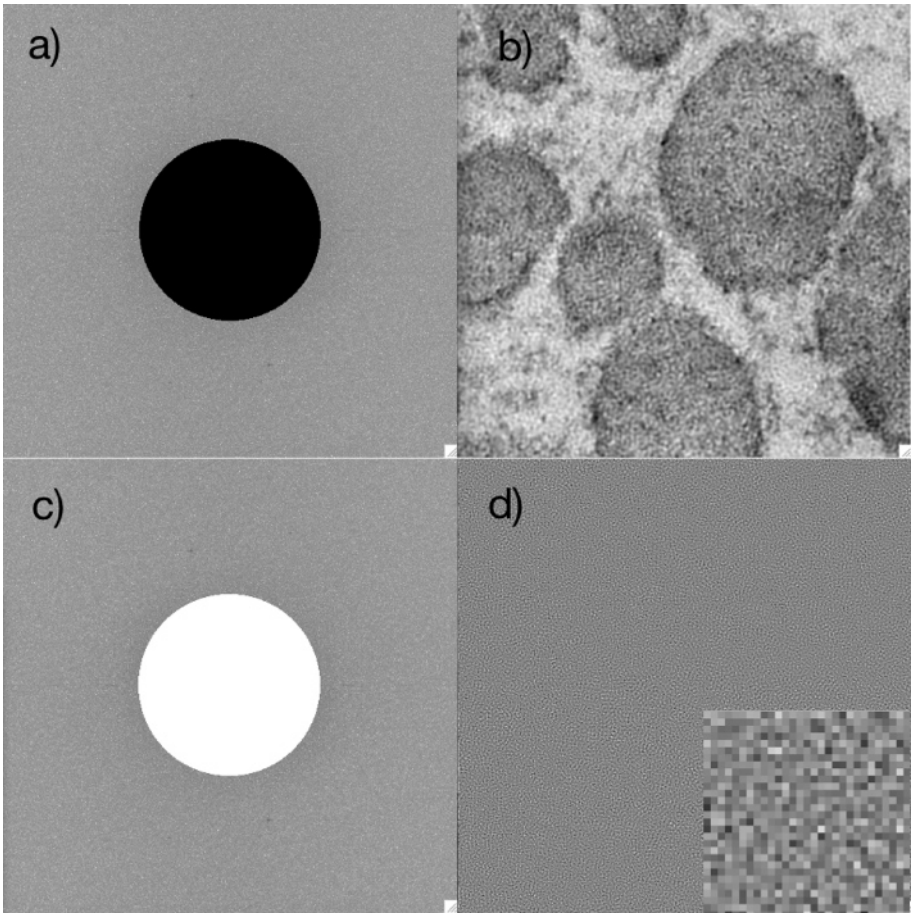


Figure 6.15: a) Low pass filter at a threshold of 5 pixels, and b) inverse FFT image of a), original without frequency components of 5 pixels and lower. c) High pass filter at a threshold of 5 pixels, and d) inverse FFT image of c), containing only frequency components up to 5 pixels. The inset in d) shows a part of d) at 6x magnification.

grey values, it might be possible to selectively remove the grey values associated with the high frequency components. Although the FFT filtering is capable of removing frequency elements irrespective of their grey value, being able to further segment the image this way could conceivably be an advantage. Fig. 6.16 shows the histograms in comparison. The peak of the low pass filtered image (a) has a broad shoulder at the lighter grey value slope, but other than that the histograms of the low and high pass filtered images are almost identical.

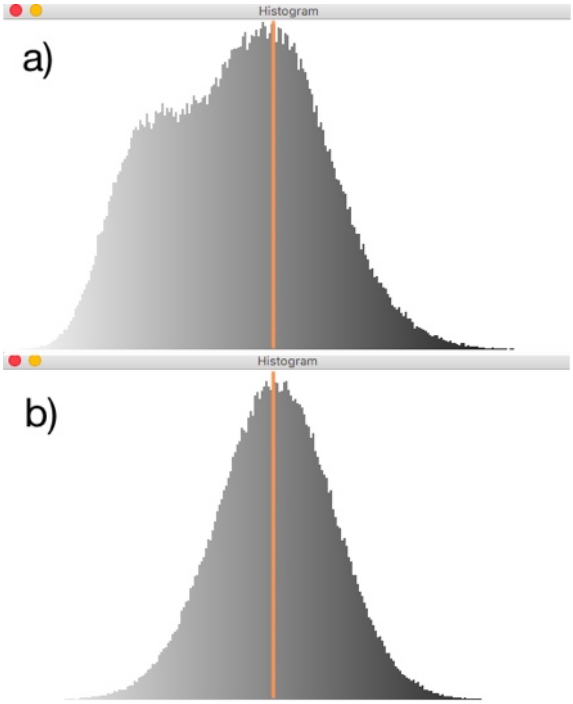
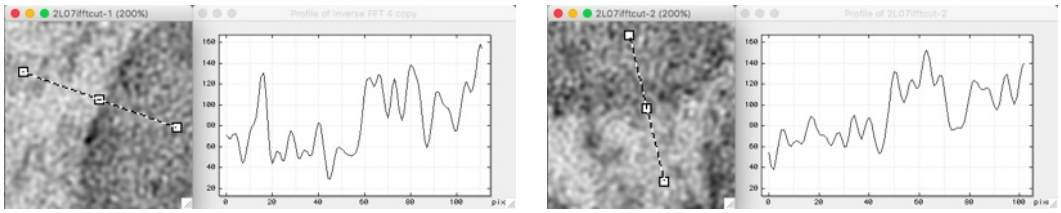


Figure 6.16: a) grey value histogram of the inverse FFT image of a low pass filter with a threshold at 5 pixels (elements larger than 5 pixels); b) grey value histogram of the inverse FFT image of a high pass filter with a threshold at 5 pixels (elements smaller than 5 pixels). The orange line indicates the peaks at identical locations on the grey value spectrum.

In order to test whether a low pass filter at 5 pixels has improved the quality of the background-blob boundary, we repeated the line ROIs shown in fig. 6.10 in the same location of clear and unclear boundaries. Fig. 6.17 shows the results of this. While the plot profiles are smoother, the peaks and troughs masking the actual edge are still present. This indicates that the noise responsible for this is at a larger scale than 5 pixels.



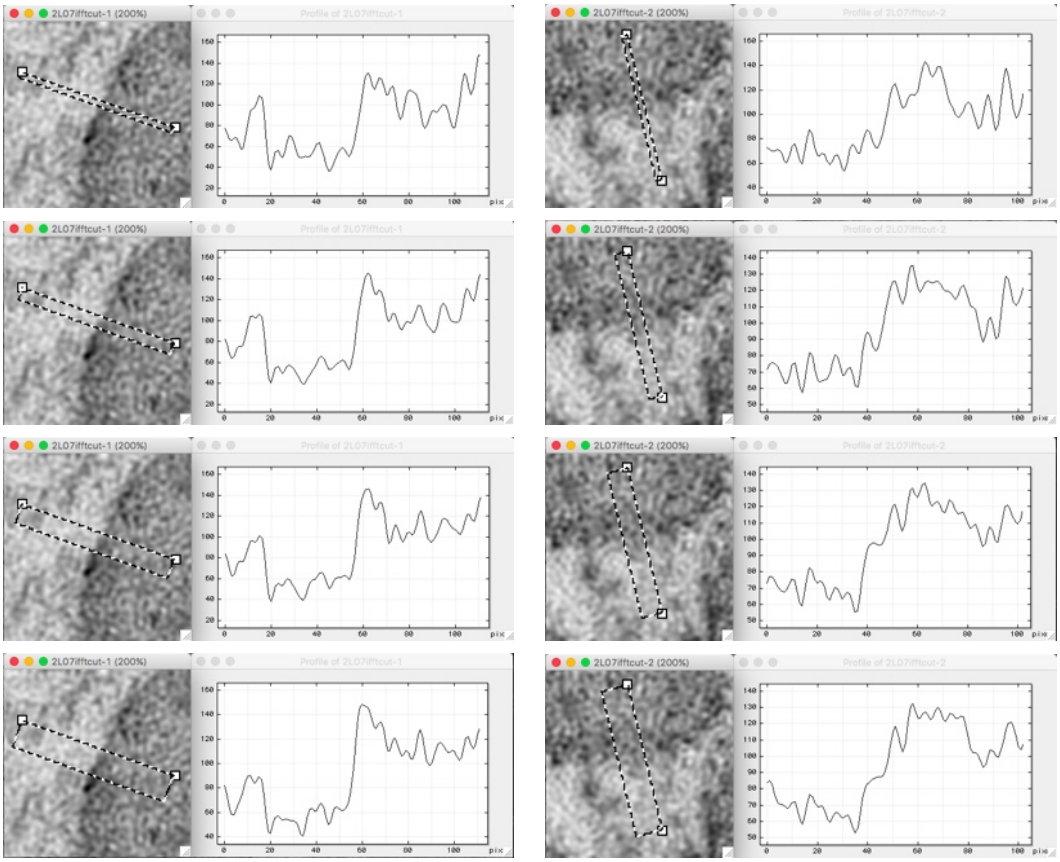


Figure 6.17: After a low pass filter at 5 pixel threshold was applied, line ROIs of various thicknesses (1, 5, 10, 15 and 20 pixels) were drawn through the same blobs as in fig. 6.10. Again, the first column shows the line ROI and plot profile over what is considered a clear boundary. Second column shows the same drawn over what is considered an unclear boundary. The plot profiles have been smoothed, but the boundaries are not more recognisable than in fig. 6.10

Since removing frequency components with less than 5 pixels has not removed the noise that prevent clear boundary identification, we extended the threshold of the filter to 10, 15 and 25 pixels. The inverse FFT images of the low pass and high pass filters at these thresholds is shown in fig. 6.18. Noise does not seem to decrease in the low pass inverse FFT images, and increases in length scale in the high pass inverse FFT images. At 25 pixels the high pass filter also starts to encompass some structures that belong to blobs, and thus eliminates information that might be useful for image segmentation.

To briefly compare the effect of low pass filtering at 5 and 25 pixels on blob boundaries, a line ROI was drawn across three blobs in both inverse FFT images. A solid line representation of the line ROI and the plot profiles of these ROIs are shown in fig. 6.19. Although the inverse FFT of the 25 pixel low pass filter shows a much smoother plot profile, it is no clearer in its blob boundaries.

Hence, it appears that noise in our image set is present continuously, and not clustered at some distinct frequencies. To confirm this, a radial distribution of an FFT image is shown in fig. 6.20. The grey values of concentric circles are averaged and plotted as a function of distance from the centre of the FFT image, or frequency. If there was more power in a particular frequency, this would reveal itself as a peak in the radial distribution.

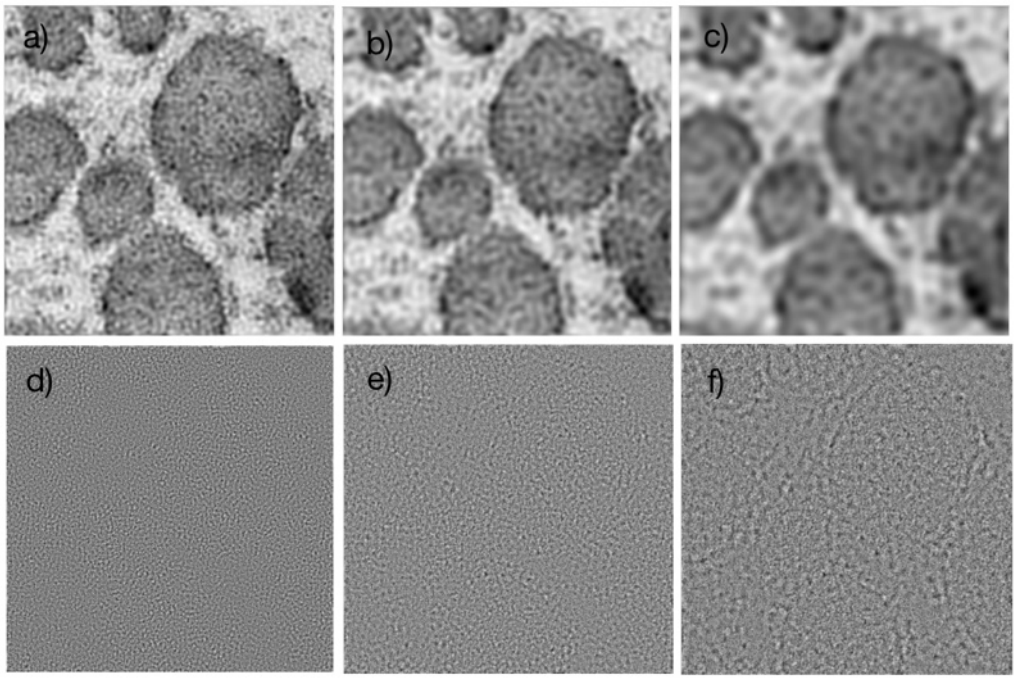


Figure 6.18: Inverse FFT images of low pass and high pass filters at various frequency thresholds. a) to c) low pass filters; d) to f) high pass filters; a) and b) threshold at 10 pixels; b) and d) threshold at 15 pixels; c) and f) threshold at 25 pixels

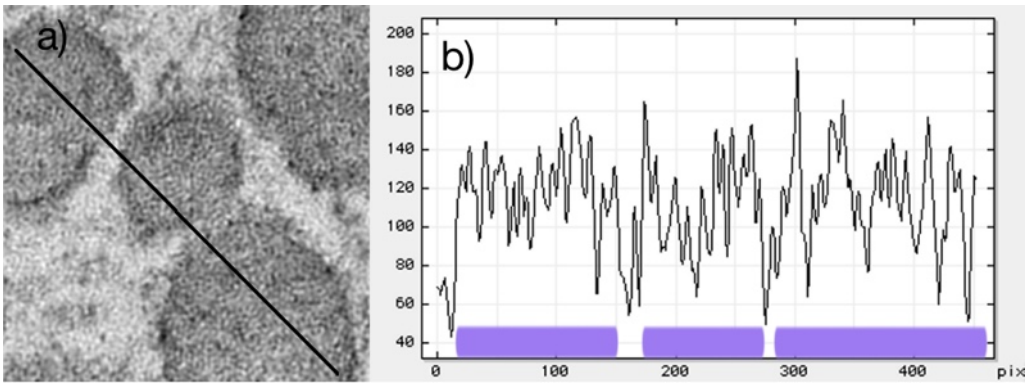


Figure 6.19: Sampling across a larger part of the image to illustrate differences between blobs and background. a) part of an inverse FFT image after a low pass filter at 5 pixel threshold, with black line indicating line ROI location; b) plot profile of line ROI in a). The areas where the line ROI crosses a blob are underlined in colour. Note the similarity in noise levels between blobs and background, and the non-obvious boundaries.

Fig. 6.20 shows a very low and shallow elevation between 0.1 and 0.2, corresponding to lengths between 5 and 10 pixels. This is further evidence that noise appears to exist on a spectrum of frequencies in these images.

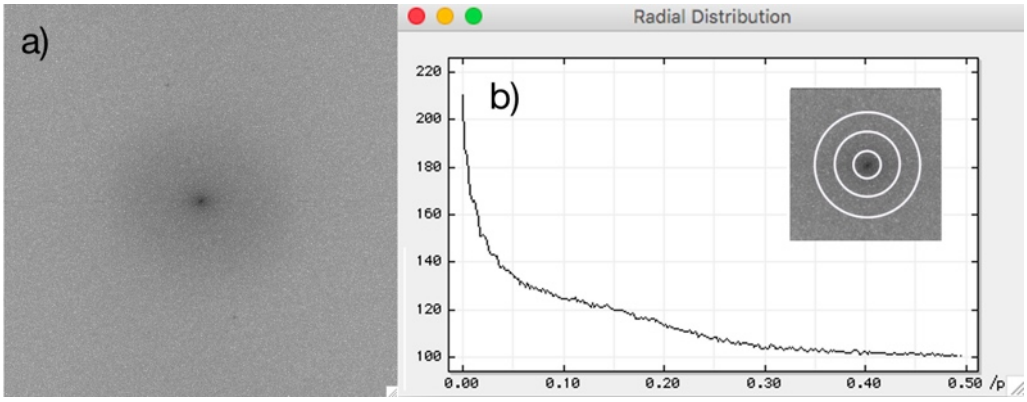


Figure 6.20: a) Example FFT image of one of our image set and b) its radial profile. The inset shows a diagrammatic representation of three image circles that, averaged, make up a radial profile plot point each. There is no peak in the radial distribution, suggesting that there is no particular, clustered noise frequency.

6.2.5 Outlines by Eye

Since the visual system is excellent at finding patterns, even where there aren't any, just looking at the blobs in an image may fool the observer into believing that there are clear structures that can be identified, when really only fragments of these are actually present. Since in the last section it was found that in both the green and red images the quality of edges varied widely, we wanted to examine to what extent edges were clear or unclear in the entire images. Thus, we traced manually the outlines of blobs over the three images, only where they were convincingly present. We did this with a clear intent to focus on the area immediately under the pen, rather than on the wider environment of the image, which could suggest to extrapolate lines without the evidence from the pixels underneath. Also, no lines were drawn where no decision could be made between more than one possible edge.

In keeping with their classification as green for figs. 6.21 and 6.22, and red for fig. 6.23, the outlines in the first two images are more complete than in the last image. Continuity of a convincing outline is one of the most important characteristics of images that are capable of being analysed.

This brief demonstration is also a useful first step to carry out when first introduced to a new set of images for analysis. The success or otherwise of tracing outlines is not only a good predictor of the difficulties ahead, but also yields some important information about the direction the analysis strategy should take for best results. For example, if the tracing results in closed objects, an analysis utilising a Particle Analysis might be appropriate, whereas if the outlines resemble lines, a texture analysis or linear Hough Transform might be a better approach.

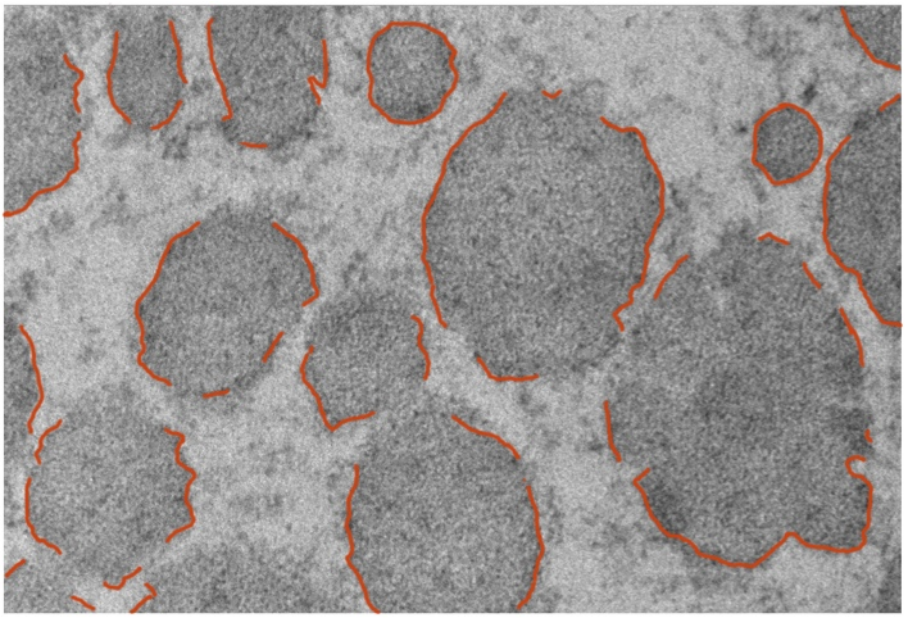


Figure 6.21: Image 2L-07: convincing blob outlines traced by hand. Most blobs have large stretches that can be traced, but few have no breaks in their outlines. Especially adjacent to other blobs lines appear to break up.

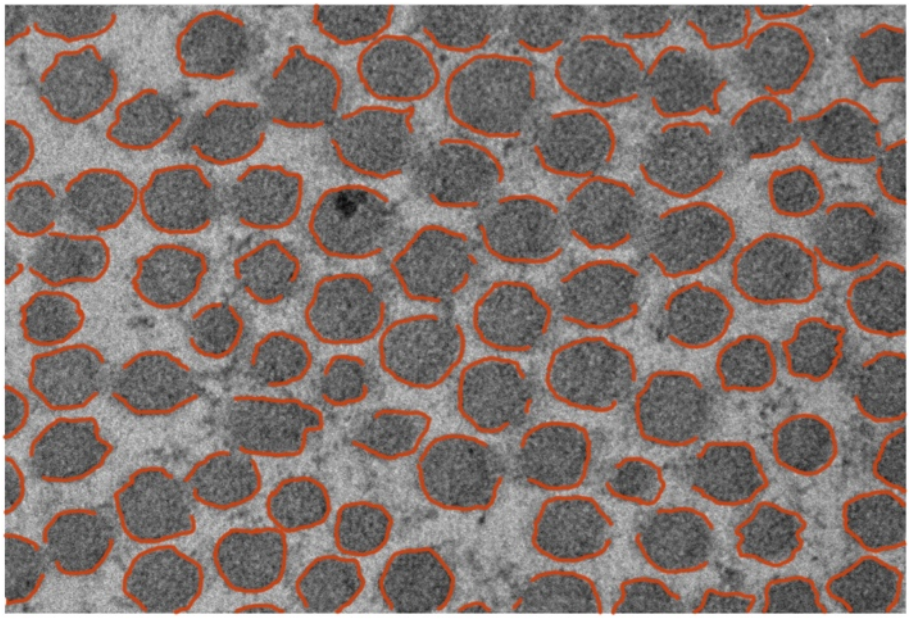


Figure 6.22: Image 2P-07: convincing blob outlines traced by hand. Here most blobs have only one gap in their outlines, but the gap does not seem to show a preference for always appearing close to neighbouring blobs. Few have an unbroken outline.

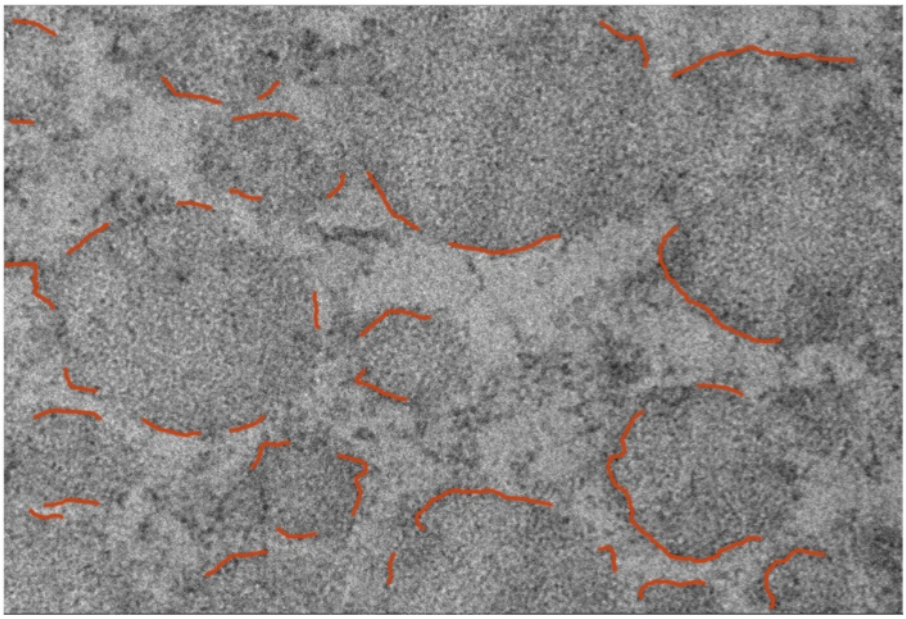


Figure 6.23: Image 2L-03: convincing blob outlines traced by hand. Very few clear outlines exist in this image. Most of the impression of blob shapes seems to be created by the visual system extrapolating to ovals. There are many places where an edge is possibly present, but could be in several locations.

6.2.6 Degrees of Image Degradation

From what has been shown so far, it appears that our set of images to some degree combine all the attributes mentioned above, which interfere with successful image segmentation. There is low contrast and no bimodal grey value distribution that might allow separating blobs from background by thresholding. Noise is prevalent to the same extent within the blobs and in the background, at a spectrum of length scales and distributed across all grey values in the image. Furthermore, the outline of almost all blobs is interrupted by stretches where the step becomes unrecognisable.

To assess whether these three factors are sufficient to thwart otherwise successful analysis methods, we generated dummy images (fig. 6.24). The first three are impaired by only one of the factors (noise (n), broken outline (bo), and low contrast (lc)), the next three combine noise and a broken outline (n + bo) and low contrast and a broken outline (bo + lc) and noise and low contrast (n + lc). The last image combines noise, broken outline and low contrast (n + bo + lc).

As exemplary analysis tools, the Hough Transform (see A.8), a Gaussian smoothing filter and a Rotating Mask (see Appendix) were chosen. In a real analysis attempt, these tools might be combined with other measures to improve the image prior to their application, and the outcome might be processed further. Also, the input parameters of these tools would need to be optimised. However, here they solely serve to demonstrate the effects of progressive image degradation, and are used at one setting each.

Fig. 6.25 shows the application of the Hough Transform to the image in fig. 6.24. The noise and broken outline alone hardly affect the ability of the Hough Transform to detect the circles, but the low contrast leads to some ambiguity (first row). When combining two

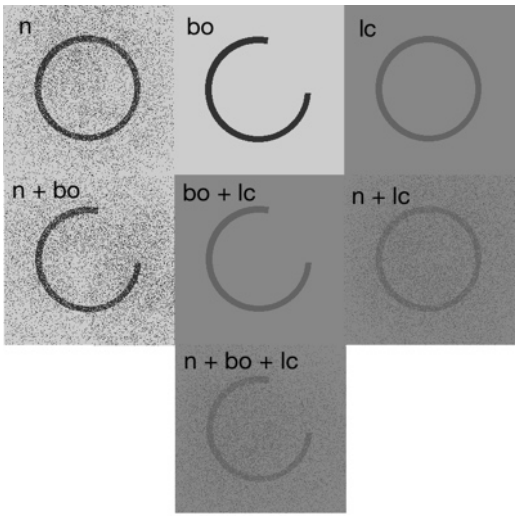


Figure 6.24: Dummy images that demonstrate progressive image degradation. The first three images contain only one aspect of degradation: noise (n), broken outline (bo), and low contrast (lc); the next three combine noise and a broken outline (n = bo) and low contrast and a broken outline (bo + lc) and noise and low contrast (n + lc). The last image combines all three, noise, broken outline and low contrast (n + bo + lc). The noise consists of single pixels at the two grey values of background and blob, unsmoothed.

degradation elements (second row), noise and a broken outline together do not significantly impair the Hough Transform, whereas the combination of broken outline and low contrast, and the combination of noise and low contrast have a similar effect to low contrast alone. It appears from this that low contrast is the most problematic aspect to deal with for the Hough Transform. When all three elements are combined, however, the Hough Transform breaks down and fails to detect the correct blob.

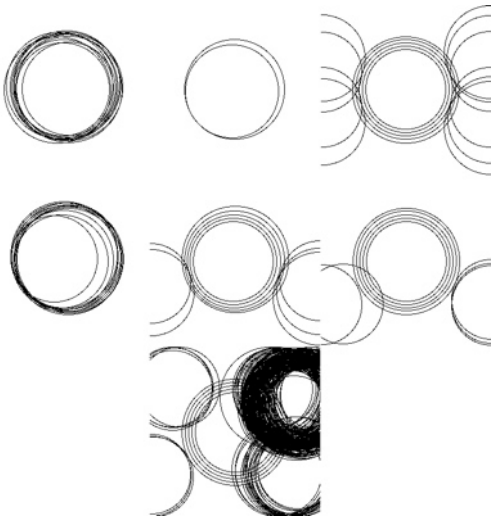


Figure 6.25: Hough Transform applied to images in fig. 6.24. For labels refer to fig. 6.24.

In fig. 6.26 a Gaussian smoothing filter with a kernel size of 5x5 pixels is applied to the image in fig. 6.24. This is intended to smooth the pixellated noise and thus aid thresholding. This kind of operation should lead to a binary image suitable for feeding into the Particle Analysis routine. Fig. 6.27 shows the image thresholded at grey value 128.

The first five images yield binary blobs that could be analysed by the Particle Analysis without problems (even though the broken outline would need to be addressed before final analysis), but the two last images, n + lc and n + bo + lc, become impossible to analyse. When these are thresholded at higher grey values the blobs themselves start to break up before the dark spots around them disappear (data not shown).

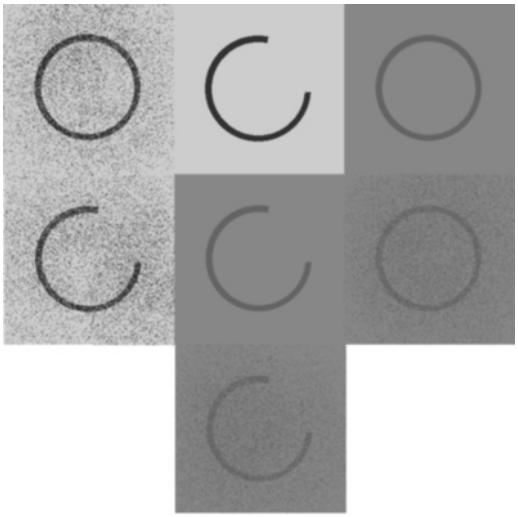


Figure 6.26: Gaussian smoothing filter applied to images in fig. 6.24. The kernel size is 5x5 pixels. For labels refer to fig. 6.24.

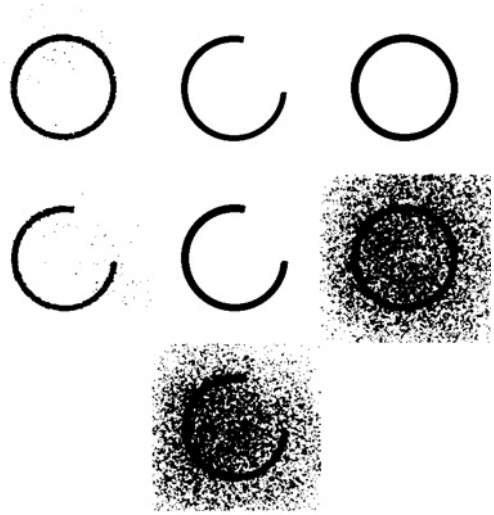


Figure 6.27: Gaussian-smoothed images from fig. 6.26 thresholded at grey value 128. For labels refer to fig. 6.24.

The Rotating Mask is an edge-preserving smoothing filter that is based on assessing the inhomogeneity in the neighbourhood of a pixel. Fig. 6.28 shows the Rotating Mask applied to the dummy images (fig. 6.24) at a neighbourhood radius of 15 pixels. This method addresses the noise in the images, but not the broken outline or contrast. Similarly to the Gaussian smoothing filter, its main purpose is to prepare blobs in such a way that the Particle Analysis is able to correctly detect and measure the blobs. The pixellated noise has been removed successfully, while a sharp blob edge has been preserved. Fig.

6.29 shows this image thresholded at a grey value of 128. Like in the case of the Gaussian smoothing, the first five images with only one or two aspects of degradation are well prepared for the Particle Analysis, but the last two images, n + lc and n + bo + lc, cannot be analysed.

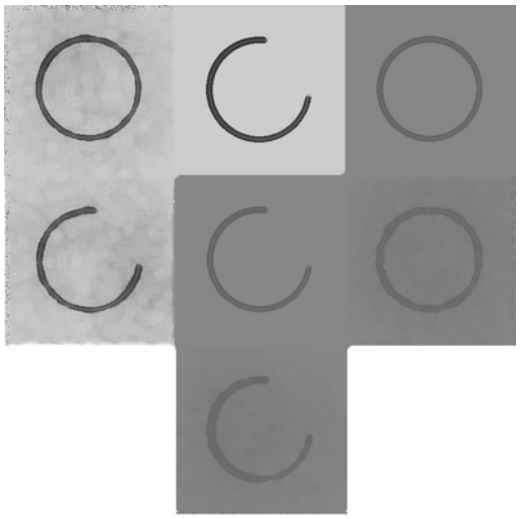


Figure 6.28: Rotating Mask applied to the image in fig. 6.24. Neighbourhood radius 15 pixels. For labels refer to fig. 6.24.

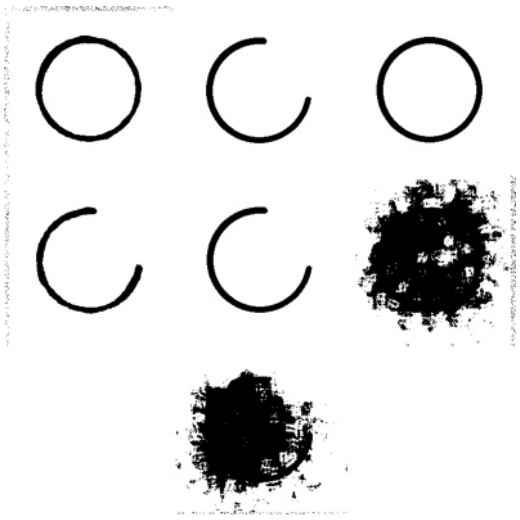


Figure 6.29: Rotating Mask image in fig. 6.28 thresholded at grey value 128. For labels refer to fig. 6.24.

The results from these dummy images suggest that a combination of low contrast and noise is the most disruptive degradation aspect in the case of the smoothing tools we used here. Furthermore, when low contrast, noise and a broken outline are combined, also the Hough Transform, which could still cope with low contrast and noise combined, broke down. It would appear that these three conditions are sufficient to prevent successful analysis. Therefore, our image set, which in many areas suffer from the combination of these three, might not be possible to analyse satisfactorily for this reason.

6.3 Quantification of Success - the Quality Evaluation Method (QEM)

We approached the analysis of our set of TEM images in many different ways, but achieved at most partial success with some of the better images, but never a completely satisfactory outcome. We were aware of the different "degrees of quality" in our images, but had no way to quantify this other than through experience and visual impression. Also, the partial success of our analysis strategies suggested that some of them might be worth pursuing further, with better quality images, but again we had no way to quantify this impression. Especially in the analysis strategies that contained many variables, such as grey values for thresholding, pixel radii, number of iterations of a procedure, etc., it would be desirable to explore the entire parameter space available. This was not tried, since the restrictions of the macro code and the time necessary to run routines from macro made it prohibitive to explore all parameter combinations. Furthermore, some interactivity between the individual image and the code may allow the parameter set to be optimised for each image. To find the optimal combination, however, again some quantification of the result of each parameter set would be necessary.

Quantification of success requires the distinction between what is a good outcome and what deviates from this. The definition of "goodness" in an analysis is not always easy to determine, since the ideal result is often not known. In images where the analysis is aimed at detecting what can be seen by eye, however, visual analysis can be used as a "gold standard" against which everything else can be measured. This is done with the understanding that visual analysis may not be completely exact or accurate, but is the best available approximation. It also should be noted that this introduces an element of subjectivity that might make the quantification done by one person different from that done by another person. For visual analysis, blobs were drawn by eye into the image where they were judged to be, and this thresholded and clear image was fed into the Particle Analysis routine.

Hence, we developed a Quality Evaluation Method (QEM) that is based on the differences between automatic and manual results of three different measurements and expressed in a single number, the Quality Evaluation Number (QEN). The measurements ought to be relevant to the biological information the analysis strategy is aimed at extracting, which in the case of this image set was chosen as the number of blobs in the image (n), the combined area covered by the blobs in percent of the entire image (a) and the average axial anisotropy of the blobs (ax). The QEN is given by

$$QEN = i\left|\frac{n_a}{n_m} - 1\right| + j\left|\frac{a_a}{a_m} - 1\right| + k\left|\frac{ax_a}{ax_m} - 1\right|$$

The subscripts a and m designate "automatic" and "manually" measured, respectively.

n = number of blobs in image

a = area of blobs in image

ax = average axial anisotropy of blobs (excluding those that touch the edges)

i, j, k are weighting factors that allow a shift of emphasis between the different pa-

rameters.

If the automatic measurements coincide completely with the manual, the terms $|(n_a/n_m) - 1|$, $|(a_a/a_m) - 1|$ and $|(ax_a/ax_m) - 1|$ all become zero. Hence, the better the success rate the closer the QEN will be to zero. Each term's absolute value increases from zero when the automatic analysis overestimates or underestimates the number compared to the visual analysis. It is therefore not possible for the QEN to become falsely zero, or close to zero, in the case where some measurements are underestimated and others are overestimated.

The measurements chosen here (number of blobs, area percent and axial anisotropy) may not be of equal importance in the analysis. Therefore, weighting factors (i, j, k) are introduced to allow for modifying their emphasis. In principle, any number of measurements could be used, in which case giving them more or less weight would be desirable. Also, if no satisfactorily low QEN can be achieved with one set of weighting factors, changing them might improve the situation.

The QEM has uses in three areas.

Firstly, in can measure the overall efficacy of an analysis strategy by comparing the outcome of the automated to the manual measurements in a number of suitable images of good quality. This way, different analysis strategies can be compared quantitatively and ranked, for a given set of images.

Secondly, when using an analysis strategy known to work well, the quality of images and sets of images can be compared, either to the average of a standard set or of manual measurements, by applying the same analysis strategy to the different sets of images. Therefore, it would be possible to set a threshold QEN for an image or a set of images that has to be met before it is included in any analysis. This could also be automated. Care must be taken, however, not to bias the results by choosing measurements that might actually distinguish the image sets. If measurements are chosen that are expected to be the same between groups, this should not be a problem. A more labour-intensive way would be to manually measure images from each set and compare the automated analysis in these. In this case it is important to chose representative images.

Thirdly, the QEM can provide valuable information when exploring the parameter space of an analysis method. Each parameter in the analysis will have an influence on the final result, but without numerical quantification the user has to inspect the outcome by looking at each image or each set of figures in turn, before deciding which combination appears best. Not only is this sometimes hard to judge, but it is also time-consuming and may miss some parameter sets. It would be preferable to at least pre-select the results that have achieved a minimum success. By first carrying out a manual measurement and then calculating the difference to this of the automatic measurements, as the analysis method cycles through the parameter combinations, the suitability of the parameter sets can be assessed as the analysis is running, and the optimal parameter set (OPS) can be found without user input at this stage. Also, a multi-dimensional parameter space can be constructed (the number of dimensions depending on the number of parameters) and existing algorithms employed to find absolute and local minima, since it is conceivable

that multiple combinations of parameters might produce a good outcome.

The duration that such an algorithm would take to run, and the complexity of the requirements, make the implementation only feasible in source code, not in macro language, which was originally designed to automate only simple routines calling inbuilt functions. Therefore, the concept of the QEM was developed here only theoretically. Nevertheless, it is hoped that these concepts may prove useful in the future in a wider area.

To demonstrate a very brief example of how the QEM can be used in its simplest form, we show a quantification of the outcome of an analysis strategy we developed, using just one parameter set. Fig. 6.30 shows the flow diagram of the analysis, which consists of an initial Rotating Mask with iterations of binary erode and dilate operations to separate the blobs while preserving their area as much as possible, followed by a Watershed, to further introduce dividing lines (for explanation of these operations see Appendix A). In this example, the parameters that could in principle be altered were diameter of the Rotating Mask, threshold grey value, "count" number for erosion and dilation and iterations for erosion and dilation, as well as the less important cut-off points for size in the Particle Analysis. Fig. 6.31 a) shows a "good" image, whereas fig. 6.31 b) shows a more challenging one. In the second row, c) and d), the images resulting from the analysis routine as fed into the Particle Analysis are shown. The third row shows the manually identified and thresholded blobs of the respective images, which are also then used for the Particle Analysis. It is clear from observation that the analysis has been more successful in c) than in d).

While in c) the blobs are roughly where they would be expected, the Watershed has not been able to separate the blobs correctly. Some blobs have been split up and will be identified as separate by the Particle Analysis. Also, the outlines are not always where one would expect them, and this will impact measurements of the axial anisotropy specifically.

In contrast, the image in d) bears little resemblance to its origin in b). Only some edges of blobs have been preserved, but will be identified as separate particles, and the overall shapes have been lost. The Particle Analysis reports the measurements for number of blobs, area percent and axial anisotropy as detailed in table 6.1.

Measurement	**2L-04** automatic	**2L-04** manual	**2L-03** automatic	**2L-03** manual
Number of blobs	33	17	73	15
Area percent	43.3	52.9	35.0	57.1
Axial anisotropy	2.08	1.68	2.31	1.63

Table 6.1: Automatic and manual measurements of numbers of blobs, area percent and axial anisotropy in the images shown in fig. 6.31.

	2L-04	**2L-03**
QEN	1.36	4.67

Table 6.2: QEN for images 2L-04 and 2L-03 when analysed by the routine outlined in fig. 6.30 using one parameter set.

When these numbers are inserted into the QEN formula, the QEN for image 2L-04 is

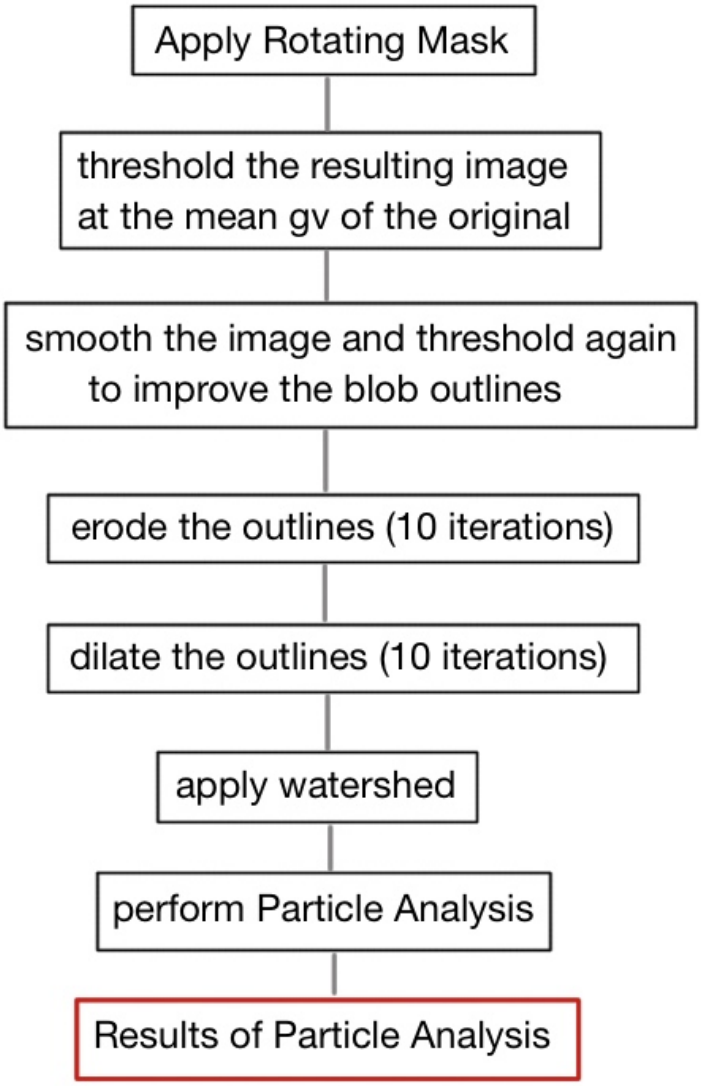


Figure 6.30: Flow diagram of example analysis strategy that achieved partial success. Here the parameters are diameter of the Rotating Mask, threshold grey value, "count" number for erosion and dilation, and iterations for erosion and dilation, and the cut-off points for size in the Particle Analysis.

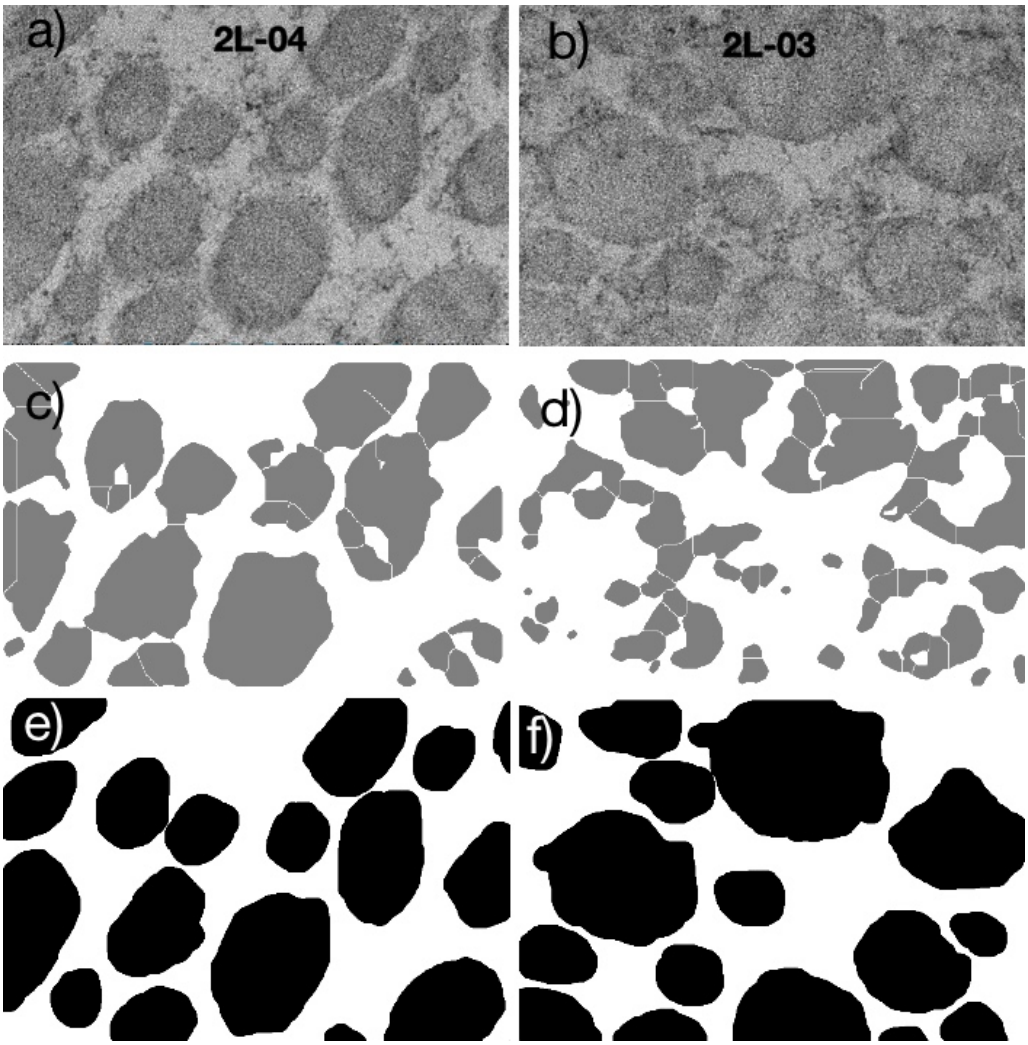


Figure 6.31: Two images for a simple illustration of the QEM. a) is a relatively good quality image (image 2L-04) and b) is more degraded (image 2L-03). The results for the images above, after the analysis routine outlined in fig. 6.30 has been applied, are shown in c) and d). In this particular case the diameter of the Rotating Mask was 31 pixels, threshold grey value was the mean grey value of the starting image, "count" number for erosion and dilation was 3 in both cases, and iterations for erosion and dilation was 10, also in both cases. The manually outlined blobs for both images are shown in e) and f). Because of the small area and for the purpose of this demonstration the particles touching the edges have been included.

1.36, and the QEN for image 2L-03 is 4.67 (table 6.2). Thus, what can be clearly seen by eye has not only been confirmed, but also quantified. This example done by hand serves merely to illustrate the principle of the QEM. The power of this method lies in the more complex uses discussed above, for which implementation in source code will be required.

6.4 Future work

Analysis of biological TEM images appears to be intrinsically problematic, if judging from the very sparse literature available. It is likely that the staining methods themselves always produce a certain level of graininess, even in good quality images, and that this noise - present in features as well as background - presents a hurdle that is not easy to overcome. Developing an analysis strategy for biological TEM images might initially be more likely to succeed if the images are of the best quality achievable. Nevertheless, even with suboptimal images, the QEM might provide a guideline for the development of successful algorithms in the grey area between complete success and utter failure that an assessment by eye could not provide.

There are now other imaging modalities that produce images of comparable or even better magnification as TEM, and so there are in many cases alternatives that might produce images easier to analyse. However, there will be cases where some samples by their nature, or by the nature of the required data, will have to be studied by TEM. Furthermore, transmission electron microscopes are probably more widely available than newer instruments, which will ensure their continued use, in spite of the drawbacks. For these reasons it would be desirable to develop an algorithm that can overcome the problems outlined above. Apart from the development of the QEM, none of the TEM work in this thesis (which has not been shown) approaches this goal in any significant degree, and hence it would be meaningless at this stage to attempt to base future work on it. However, repeating the analysis strategies attempted before with higher contrast, less noisy images, and using the QEM when assessing the results, might yield some pointers as to a promising road ahead.

The most important outcome of the TEM work is the quantification method of the QEM, and while it might not always be applicable (for instance where there is no “gold standard” to compare with), it is a very useful tool where it can be applied. Using this method of quantification it is possible to compare analysis strategies in their final, as well as in their development stages, in a meaningful and near-objective way. The benefit of the QEM is relevant to a very wide range of image analysis problems.

The most practical application of the QEM is in the automatic exploration of the entire parameter space of an analysis routine. It requires a limited user input of manual identification of image characteristics in a few images. In most cases, it is likely that the images in a set are similar enough that the OPS found for a small number of images can also be used in an at least near-optimal way for the rest of the set.

The QEM might provide new information. If the OPS could be found easily for each image and varied between images, the distribution of the OPS within a group of images might be linked to other image characteristics that have some influence on the parameters.

Understanding these links might lead to a better understanding of the images themselves.

As already mentioned, when using an analysis known to be working with good efficiency, images could be categorised according to their QEN number. Especially when using large numbers of images, however, a measurement standard that does not require too much user input would be desirable. This might be an easily identifiable aspect of the image, which can be assumed to be constant between images, and might be unrelated to the characteristics targeted by the analysis. If data from images with a spread of known QEN numbers were pooled for analysis, their weighting could be adjusted, since images whose measurements of one quantity diverge further from a standard than the mean are also likely to have a greater error in the measurements of other quantities, and their data are therefore less reliable.

As already mentioned, the QEM requires some form of standard against which to measure the outcome of an analysis. Additional ways of providing a standard that does not require manual measurement would be helpful for situations where manually measuring a few images would not suffice. The nature of such standards, however, depends very much on the image set under investigation and would need to be considered for each case anew. Further work is also needed to find suitable algorithms to allow identification of local minima in the multidimensional parameter space, which would allow the analysis method to do much of the optimisation inside the code. More questions and areas for improvements might open up when this work is actually initiated, but at this stage the breadth of opportunities of the QEM merely points to many avenues of development.

6.5 Conclusion

In spite of numerous attempts, employing a wide range of strategies, no robust or widely applicable algorithm could be developed for the analysis of the set of TEM images at our disposal. The images are characterised thoroughly and the reason for the difficulty in analysis is identified through recreating the same effects in dummy images. A novel numerical method for quantifying success of the analysis method is presented, which has near-universal applicability for the development of image analysis algorithms in general, and may also be used for numerical assessment of image quality.

Chapter 7

Bonus Chapter

Image analysis is primarily concerned with the analysis of data that occur in a certain form (a range of integer values associated with x and y coordinates). The content of such data (what they represent) is important in guiding the aims of the analysis, but is only of secondary importance when it comes to choosing which analysis tools and methods to apply. Hence, we have argued that our analysis strategies can be applied to other images unrelated in content, but similar in structure.

In this chapter we apply the two macros developed for the analysis of collagen images obtained by AFM and PLM to images that are unrelated in content to the collagen images, but similar in structure. (All images were converted to 8 bit grey value images for analysis, although the originals are shown here.)

The results of these "experiments" carry no claim to any scientific significance, since disparate, single images are used, and we lack the background knowledge with regard to which characteristics might be of interest to extract from these images. Nevertheless, such a play-around might open the door to new ideas and demonstrates the point of the wider applicability of an image analysis strategy.

7.1 Play with the PLM Analysis

As a reminder, the quantities *AxanisAve*, *AxanisDev* and *AngleR* used in the PLM analysis are summarised in Chapter 4. Since the graphs presented below are from only one image, and not a group, there are no error bars on the data points. The images were chosen purely on the criteria of whether their structure had any similarity to the images of the PLM and SHG microscopies.

7.1.1 Millennium 2 Simulation

The Millennium Simulations ([140]) attempt to model the formation and evolution of large-scale structures in the universe (galaxies and bigger) by creating a large segment of an initially very smooth universe (dominated by "Dark Matter") and allowing it to evolve *in silico* under basic physical laws. These are very large and challenging simulations, but they can be related to the "real" universe by comparing the patterns of large scale structures formed. Those structures evolve with time, starting out smooth and gradually becoming more clumpy under the influence of gravity, with a network of complicated filaments and large voids. These features are clearly seen in 2-dimensional images of the 3-dimensional

simulation and so useful information may be gained about the statistical properties of the structure and its temporal evolution by image analysis. The images in fig. 7.1 are snapshots of the continuous evolution of the simulation.

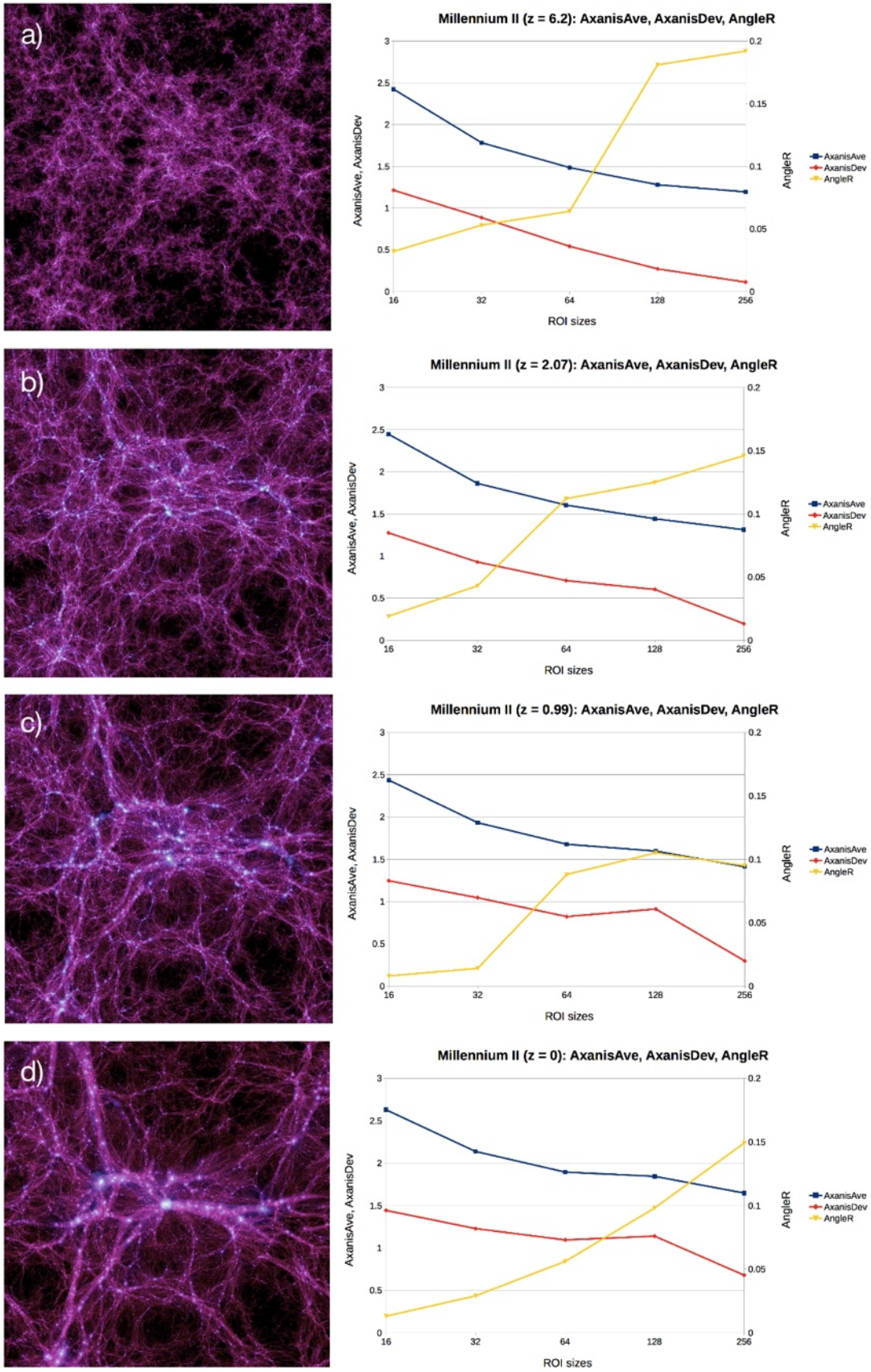
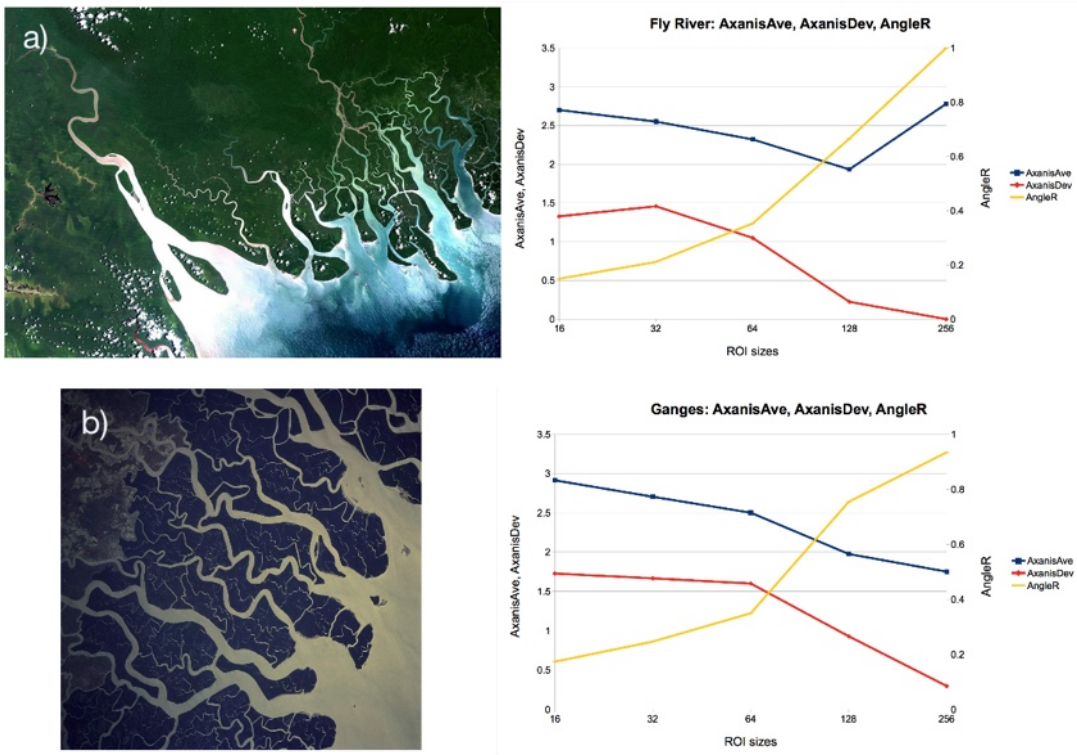


Figure 7.1: Images of the Millennium 2 Simulation of the formation of structure in the universe at different red shifts, with their "analysis results". These images show a 2D slice of the original 3D data. The scale is 100 Mpc/h each side, the image size is 1024x1024 pixels. a) red shift z = 6.2; b) z = 2.07; c) z = 0.99; d) z = 0; larger red shifts denote a universe from an earlier time, thus z = 0 represents the present day

The curves for the *AxanisAve* and *AxanisDev* hardly change at all at different ages of the universe, which is perhaps surprising, but is likely related to an element of uniformity at all scales in the modelling of the universe. The curve for *AngleR*, on the contrary, varies in an interesting way with the age of the universe in the model. At the smallest ROI size there is little difference, but as ROI size increases, the behaviour changes. In a young universe (z = 6.2) the *AngleR* increases quite steeply between ROI size 64 and 128 to about 0.2 at ROI size 256. As the universe ages (z = 2.07 and z = 0.99) the increase happens between ROI size 32 and 64 and only reaches 0.15 and 0.1, respectively, at ROI size 256. In the present day universe (z = 0) the increase in *AngleR* with ROI size is uniform and also only reaches 0.15. Since *AngleR* is an indication for how widely the angles of the ACF particles are spread (with *AngleR* = 1 being parallel), this in conjunction with the unchanging curves for the *AxanisAve* and *AxanisDev* is an interesting "result". Without in-depth knowledge of the simulation, however, it is impossible to interpret these graphs meaningfully.

7.1.2 River Deltas

Since river deltas often show similar structures at different length scales, it suggested itself to apply the PLM macro to satellite images of river deltas (fig. 7.2. We obtained the images from the internet, and hence their scale and image size are not consistent.



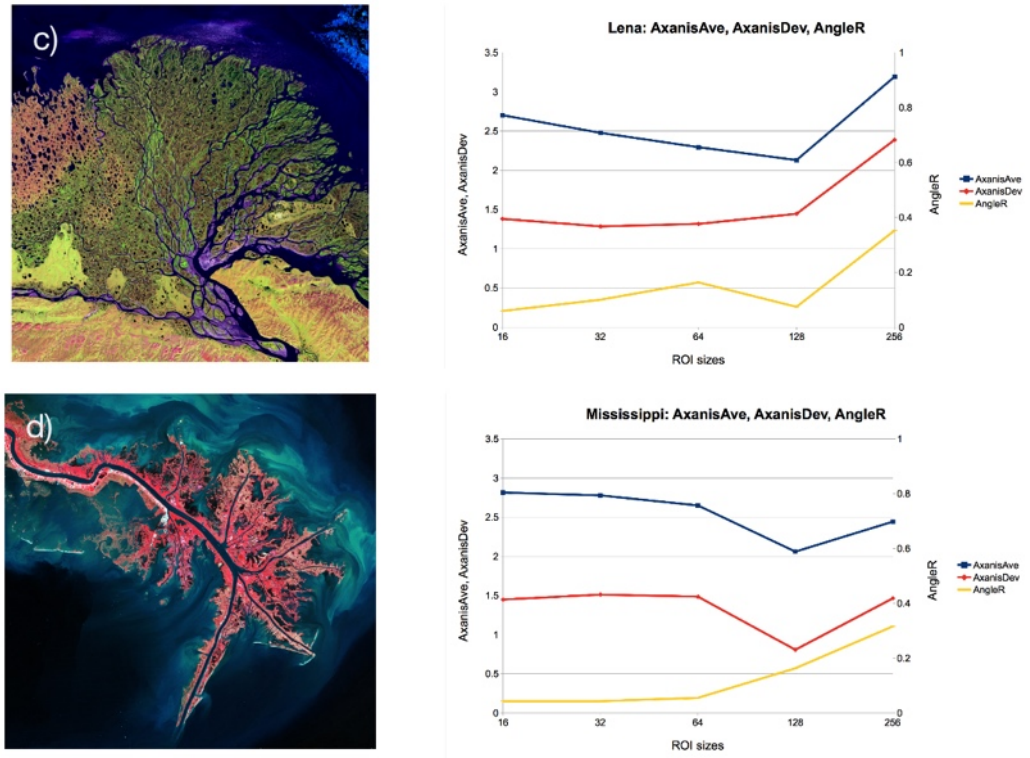


Figure 7.2: Satellite images of river deltas, with their "analysis results". Image sizes around 600 pixels. a) Hawooi And Kikori Deltas (Fly River) in Papua New Guinea (credit: NASA); b) Ganges in India and Bangladesh (credit: NASA); c) Lena in North-Eastern Siberia (credit: NASA) and d) Mississippi in Louisiana (credit: ESA)

The first noticeable peculiarity of these graphs is that for the Fly River and Ganges the curves for the *AxanisAve* and *AxanisDev* slope down at large ROI sizes, whereas for the Lena and Mississippi they dip and then slope upwards again. This is probably related to the fact that, when considering a ROI size of about 1/6 of the whole image, there is more uniformity in the images of the Fly River and Ganges, whereas in the images of the Lena and Mississippi there is more variation at this scale. The *AngleR* for the Mississippi is very low at all scales, indicating a wide range of directionality in this image, probably caused at small scales by intensity variations in the body of water, and on large scales by the uneven distribution of structures (land and water) in the image. This is a reminder that the analysis takes into account whatever is in the image section, not just the feature that the eye is drawn to, as in this case the meandering of the arms of the delta. For example, in the case of the Mississippi the colouration in the sea is going to influence the ACF as well as the prominent land (red) surrounding the river. Hence, for the serious comparison of river deltas images should not contain significant proportions of land or sea, or should be thresholded before analysis.

7.1.3 Trees

When looking for structure on different scales found in the same location, trees are probably the first thing that meets the eye. Especially in winter when the absence of leaves reveals the underlying structure, the progression of large branches, with their thicknesses

and directions, to smaller and smaller twigs and twiglets, which point in their own directions, invites their analysis along PLM lines (fig. 7.3).

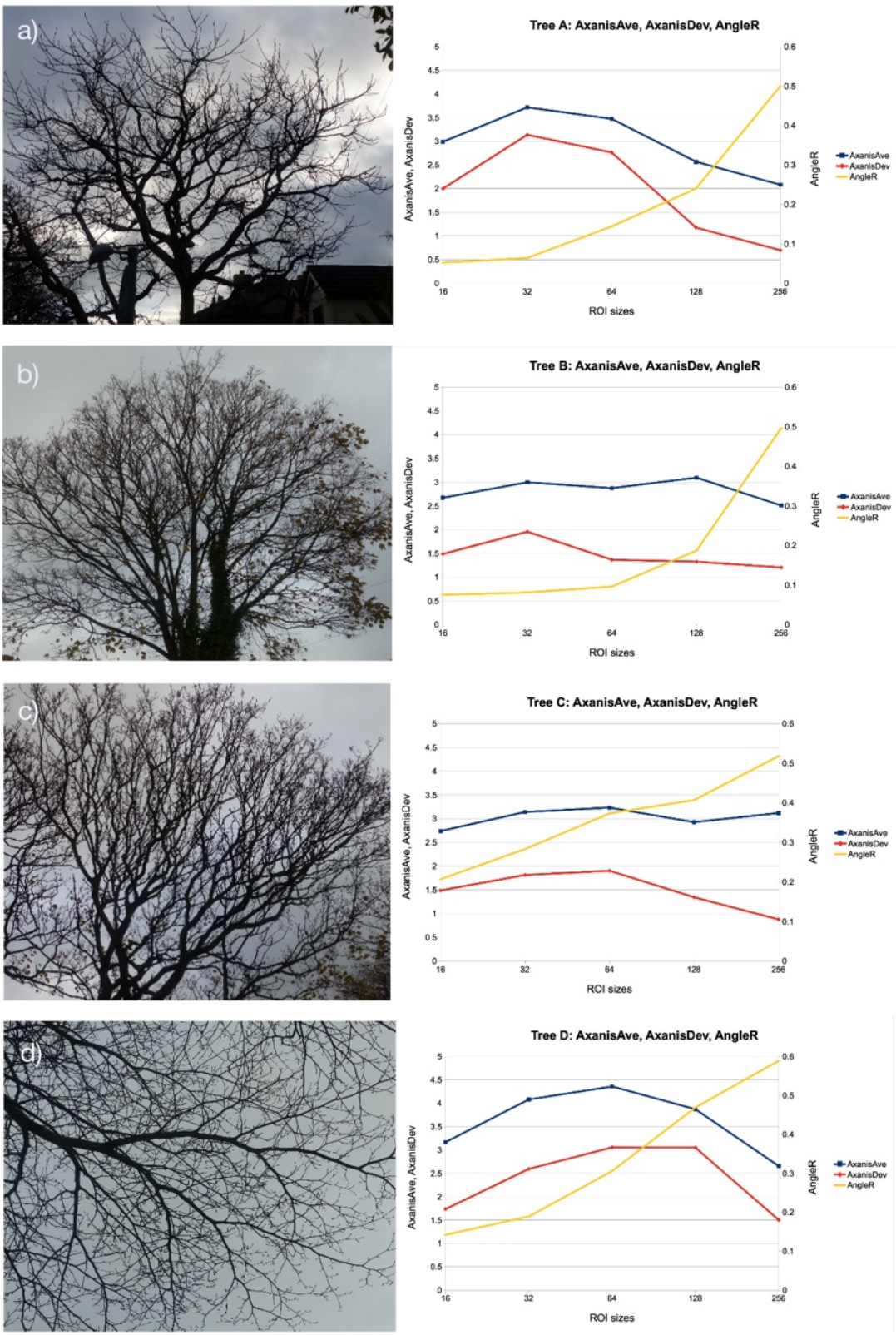


Figure 7.3: Three images of trees (a - c) and part of a tree with fractal-like properties (d). Image sizes are on the scale of about 1000 pixels. Since a tree is a 3D object, some of its structure is obscured, and therefore the results will to a certain extent depend on the position of the photographer.

The graphs in a) and d) do not differ substantially in general tendencies, and the same is true for the graphs in b) and c). The *AxanisAve* (and *AxanisDev*) slope downwards from ROI size 64 up for a) and d), indicating a greater degree of self-similarity in ROIs of larger size compared to the smaller ones. In d) the *AxanisAve* reaches a maximum at ROI size 64, suggesting that this size encompasses the twigs with highest variability in direction. In b) and c) the *AxanisAve* stays almost the same over all ROI sizes, and the *AxanisDev* changes only slightly. When examining images b) and c) by eye, it seems that the tree structure and the direction of the branches is quite similar at a range of length scales, which would lead to similar shapes of the ACF particles at different ROI sizes. In a) and d) the smaller branches seem to diverge more. The *AxanisDev* is relatively low in comparison to the *AxanisAve* in b) and c), suggesting that the axial anisotropies of the ACF particles all over the image are relatively similar. Most of the image area is taken up by the tree structures here, but the same caveat applies as already mentioned in the section on river deltas, that all image content will be analysed, and that for instance clouds behind the trees could alter the outcome noticeably.

7.1.4 Walls

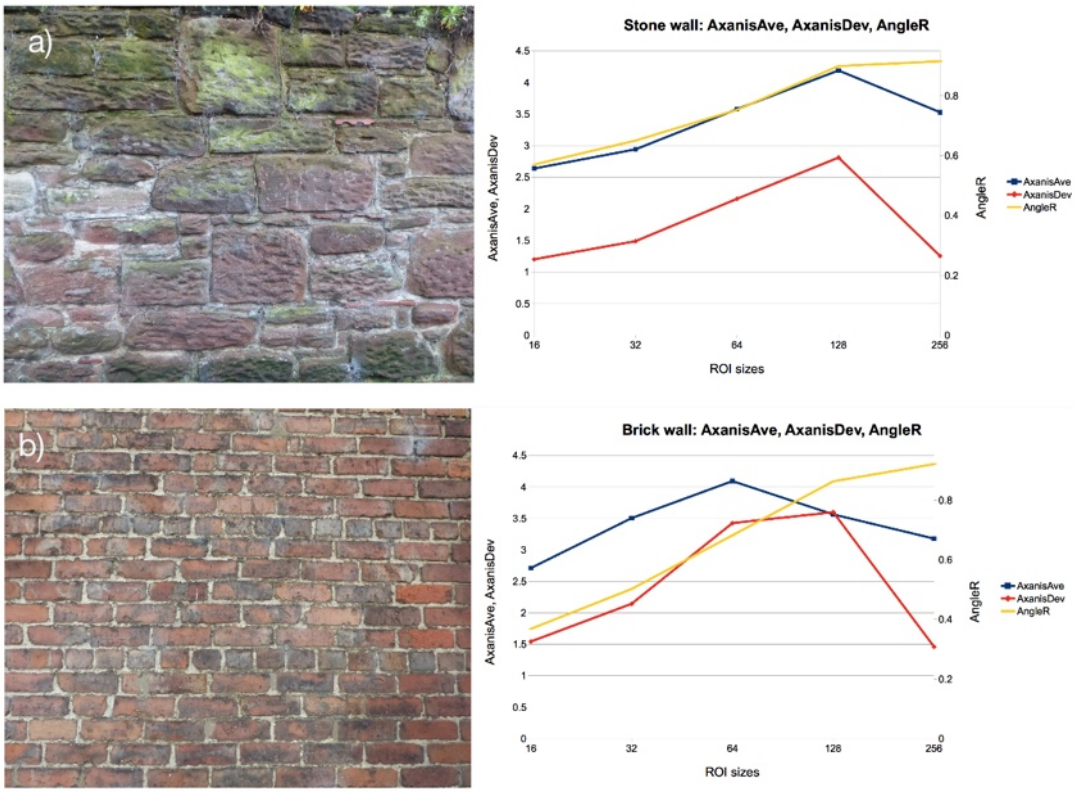
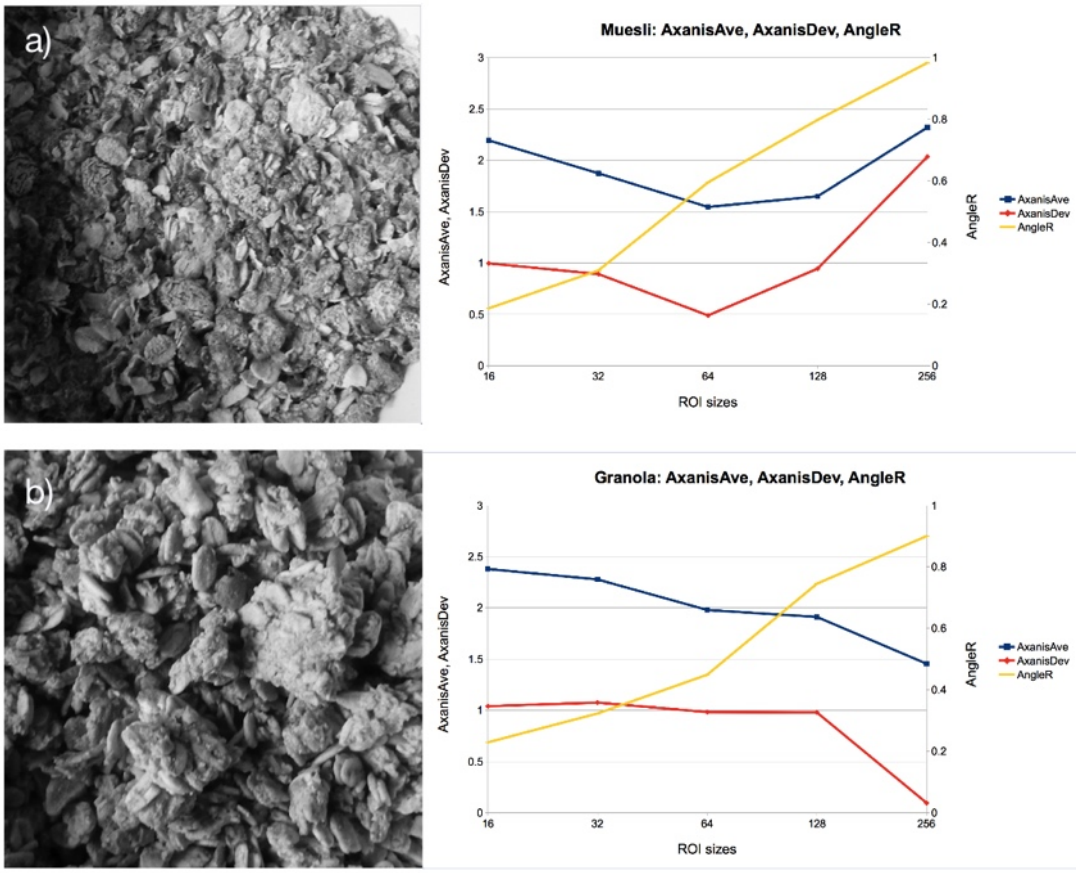


Figure 7.4: Images of an irregular stone (a) and a more regular brick wall (b). Image size for both 1280x960 pixels.

It is surprising that the graph for the more irregular stone wall should look more regular than that for the at first sight more regular brick wall (fig. 7.4). However, the main difference that affects the overall appearance is the *AxanisDev* at ROI size 128 in the brick wall graph. There the *AxanisDev* becomes as high as the *AxanisAve*, whereas it stays at the same distance to the *AxanisAve* in the stone wall graph. 128 pixels is also the width of the bricks in the image, and ROI size and feature size coinciding so well will result in some stark differences in the ACF, as the ROIs move over the image in half ROI size steps. One ROI will contain mostly brick, the next mostly border cement, etc., thus increasing the standard deviation of the axial anisotropies. This effect is not seen in the irregular stone wall. The *AngleR* starts at a higher point at small ROI size in the stone wall graph than in the brick wall graph. On close inspection, it can be seen that each stone in the wall is covered in chisel marks, all going in more or less the same direction. This small scale structure, that is missing in the brick wall, is likely responsible for the higher *AngleR* in the stone wall graph.

7.1.5 Breakfast Cereals



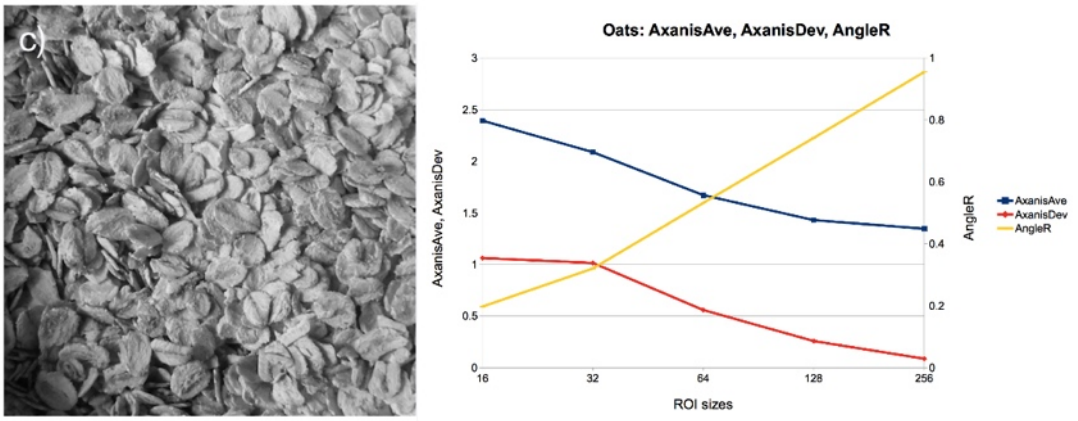
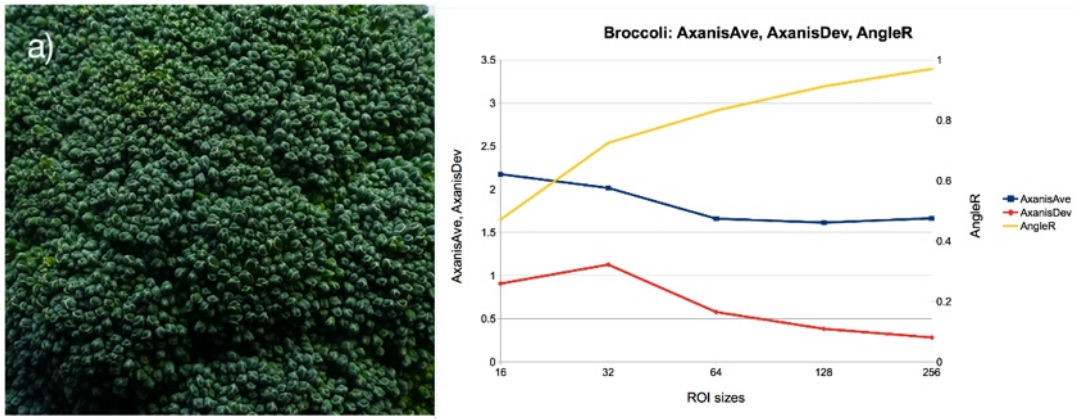


Figure 7.5: Three types of breakfast cereal: a) muesli; b) granola; c) jumbo oats. Image size 512x512 pixels.

The images of cereals were taken by hand and not under controlled distance, lighting or levelling conditions, although an attempt was made to keep these similar (fig. 7.5). It is interesting that the *AngleR* is almost identical for all three cereals. This suggests that the spread of directions of the ACF particles increases similarly at increasing length scales. The *AxanisAve* and *AxanisDev*, are also similar, except that the values at ROI size 256 are much higher in the case of the muesli. It is likely that this is an effect caused by the prominent shadow on the left of the image and the two areas of bowl in the top and bottom right hand corners. If ROI size 256 is not taken into account, the graphs differ little from each other, reflecting the similar structures of these breakfast cereals.

7.1.6 Vegetables



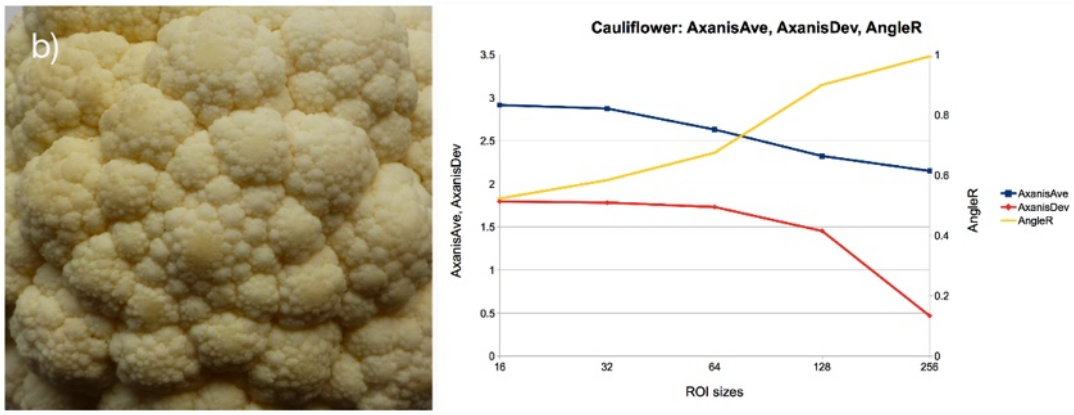


Figure 7.6: Two kinds of vegetables with structure on different scales. a) broccoli; b) cauliflower. Image sizes 512x512 pixels.

Cauliflower and broccoli (fig. 7.6) can differ considerably in their structure when going from the millimetre to the scale of about 10 cm (but not all do). Whereas the broccoli we obtained had very pronounced flower buds about 1 to 2 mm across, but little outside structure betraying the underlying floral shoots, the head of cauliflower had a very pronounced Fibonacci-like pattern of florets on the surface, on the scale of several centimetres, but the small scale patterns were flatter and thus had less contrast. The broccoli buds were about 15 to 20 pixels across, the cauliflower large scale florets around 100 to 120 pixels. Accordingly, the graph for broccoli shows little change in the *AxanisAve* and *AxanisDev*, and the *AngleR* has an upward slope at a small ROI size. The graph for cauliflower shows a higher *AxanisAve* and *AxanisDev* at all ROI sizes, indicating a greater variation in ACF particles, as would be expected, and the *AngleR* slopes upwards at a ROI size larger than the size of the florets.

The results of the "analysis" of these images, from contexts entirely unrelated to collagen, show that image characteristics can be described using the same parameters no matter what the images show, as long as the image structure has similar elements. The interpretation, however, as already discussed in the relevant chapters, is not straightforward, since many factors influence the *AxanisAve*, *AxanisDev* and *AngleR*, and a systematic use of the analysis is probably reliant on the availability of a larger number of images.

7.2 Play with the AFM Analysis

The AFM analysis requires user input by drawing a line along a structure with a regular repeat along this line, and a boundary on both sides parallel to it. The results return the width of the "fibre" along which the line has been drawn and the periodicity of the "banding" repeats. Again, images unrelated to collagen were found by looking for similar structures.

7.2.1 Fence

A simple fence in front of a wall of wood planks provides a pattern similar to a banded collagen fibre (fig. 7.7). Fig. 7.8 shows a section converted to 8 bit grey scale and inverted for ease of boundary recognition. The line ROI is manually drawn along the middle of a plank, and crossing the metal fence loops, and analysed with the AFM macro. This arrangement allows the macro to detect the plank edges as "fibre edge" and the distance of the fence parts as "banding". Fig. 7.8 a) shows the results image, the green dots indicating where the macro has found the plank edges. The plot profile along the line ROI illustrates in b) how the pattern of rust and paint on the fence in this 1 D section partly disguises the periodicity that seems clear from the 2 D image. (As already mentioned in the discussion of the AFM macro in subsection 3.3.8, the plot profile is extended to the next higher power of 2 by mirroring the profile. This is necessary for the FFT of the profile.) The power spectrum is shown in c), see figure legend.

Figure 7.7: Metal fence with regular spacing in front of wall with wooden planks at approximately right angles to the fence.

Fig. 7.9 shows a different section from the image in fig. 7.7. The line ROI is drawn in a vertical direction here, so that the spacing between the fence loops is measured as the "fibre width", and the plank width becomes the "banding". The edge points are all found correctly (a), and the plot profile shows the plank edges clearly (b). The power spectrum (c) shows many strong frequencies, which is probably due to the intensity variations in the planks.

7.2.2 Roof

Fig. 7.10 a) shows a section of a blurred photograph of a roof after analysis with the AFM macro. The green dots show correct identification of the tile width as "fibre edges". The plot profile (b) shows the tile boundaries and the power spectrum (c) the frequencies present.

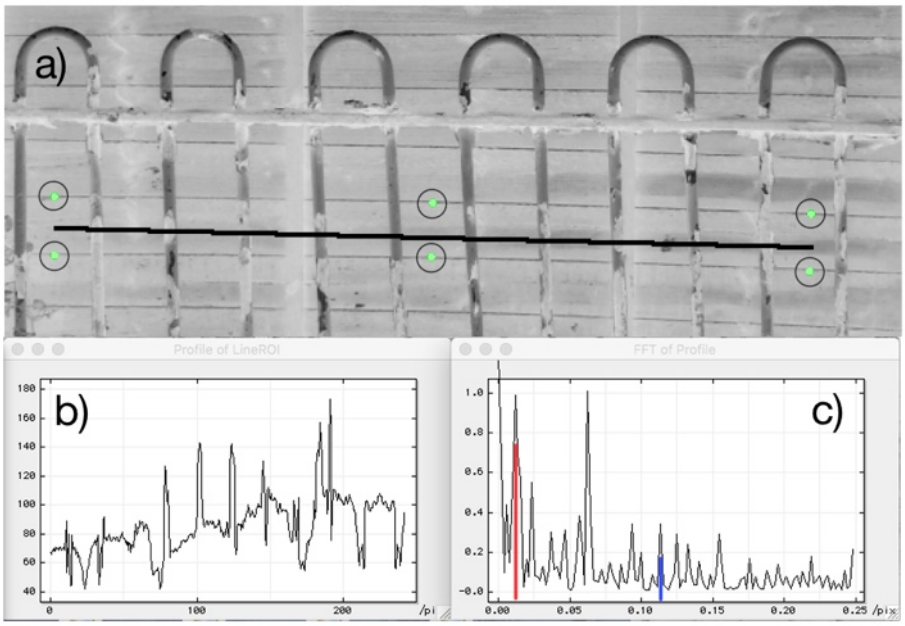


Figure 7.8: AFM analysis of a section of fig. 7.7. Note that the image has been turned into grey scale and then inverted, since the macro was developed to find dark edges rather than light. In this case the line ROI has been drawn horizontally, so that the "fibre width" result measures the width of the wooden planks, correctly at 35 pixels (not shown). a) image showing the location of the line ROI and the green dots drawn where an edge is detected (circled in black); b) plot profile along the line ROI; c) Fourier spectrum of the plot profile: the red peak at 0.022/pix corresponds to the largest repeat of fence grates at 45 pixels, the blue peak at 0.125/pix corresponds to the width of the metal bars at 8 pixels. The large peak at 0.0625/pix corresponds to 16 pixel and is likely a harmonic.

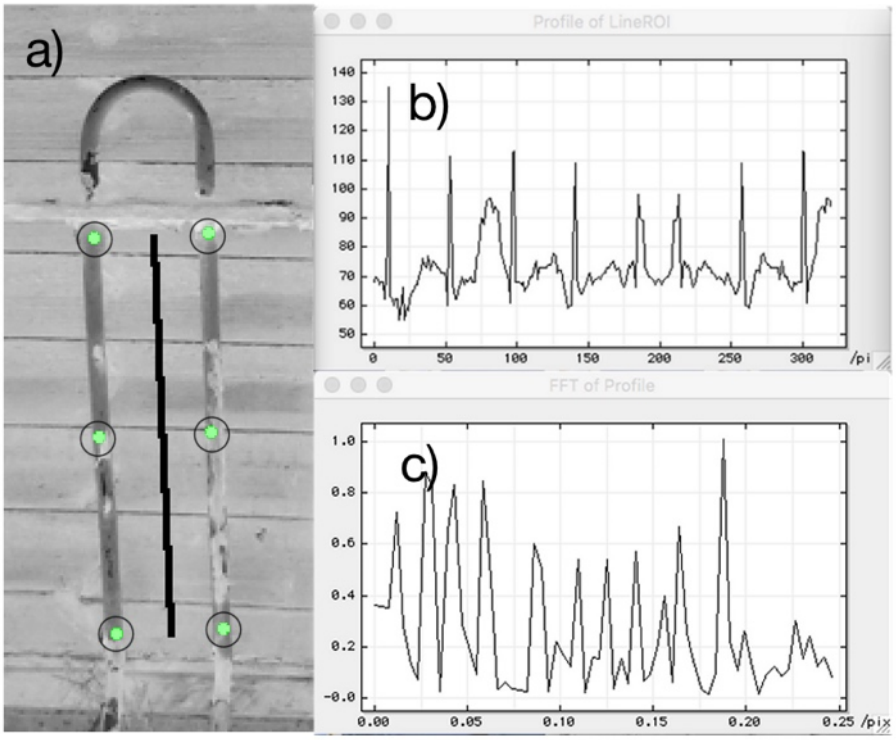


Figure 7.9: AFM analysis of a different section of fig. 7.7. This time the line ROI is drawn vertically, so that the "fibre width" becomes the spacing of the metal bars, which is again measured correctly at 45 pixels. a) image showing the location of the line ROI and the green dots (circled in black) where an edge is detected; b) plot profile along the line ROI; c) Fourier spectrum of the plot profile: There is a peak at 0.029/pix (second from left), which would correspond to the plank width of 35 pixels, but there are also other equally strong peaks, probably due to the many variations in grey value in the planks along the ROI.

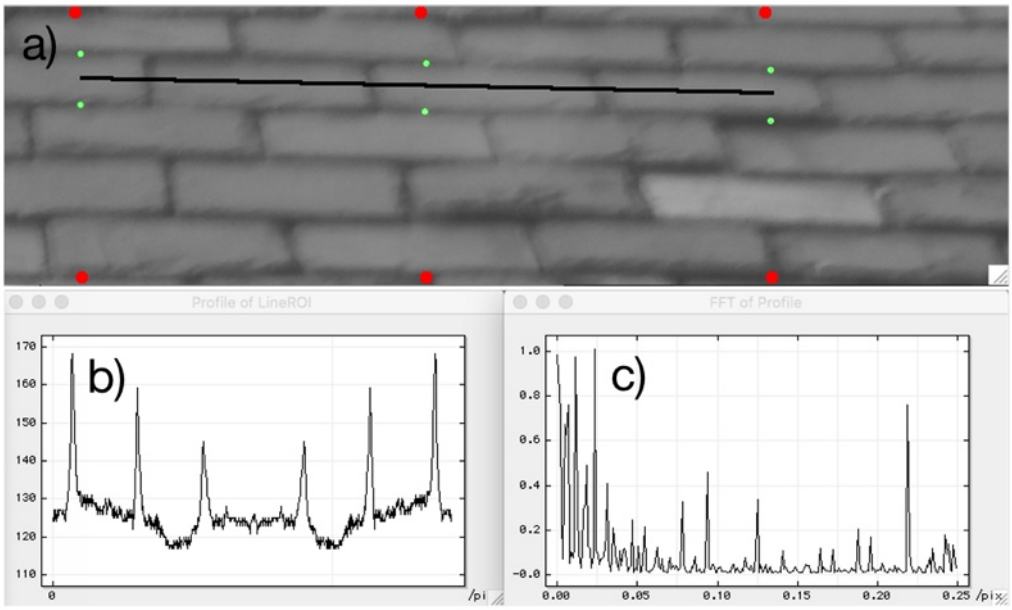


Figure 7.10: Old roof tiles of a terraced house. Although the blur was not intentional, it may have helped by smoothing out irregularities that could interfere with correct "fibre width" measurement. a) results image showing the location of the line ROI and the detected edges; b) plot profile of the line ROI; c) Fourier spectrum of the plot profile: Since the line ROI only covers two tile lengths (each tile is approximately 175 pixels long, corresponding to about 0.006/pixel), this periodicity is not clear in the Fourier spectrum.

7.2.3 Wall

A section of the image already used in fig. 7.4 b) was also used for AFM analysis. Again, the grey values are inverted for ease of edge finding. The line ROI is drawn along the long sides of the bricks. Fig. 7.11 a) shows the location of the line ROI and the correctly found brick height in the centre and on the right, as well as incorrectly identified edges on the left. The plot profile in b) shows the spacing of the bricks clearly, and the power spectrum in c) shows a strong peak identified with the dark pixels of the cement.

In spite of the relative dissimilarity between AFM scans and photographs of fences, roofs and walls, the AFM macro is very successful at finding edges parallel to the line ROI and identifying periodic structures along it. This enforces the point already discussed earlier that image content is less important than image structure for a successful analysis.

Although none of the graphs presented in this bonus chapter claims any real significance for the kind of images the macro has been applied to, this little exercise nevertheless demonstrates three things.

Firstly, it shows the robustness of the analysis strategy and its implementation in the macro. If either the concept or the realisation did not have this robustness, using different images from the set the macro was developed for might derail the analysis.

Secondly, as already stated, it shows that image content does not determine the suitability or feasibility of an analysis strategy as much as image structure does.

Thirdly, it demonstrates that image analysis in particular lends itself to interdisciplinary working. An analysis strategy that was developed in one field might easily be applied to

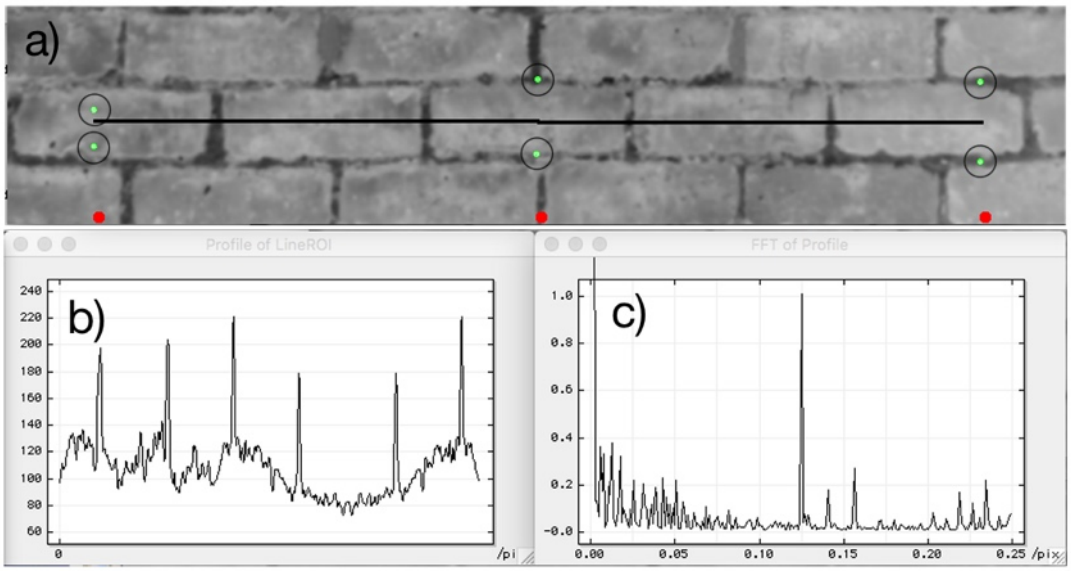


Figure 7.11: A brick wall (grey values inverted to facilitate edge finding). a) results image with line ROI location and found edges indicated by green dots (circled black for clarity), four of the six edge points coincide correctly with the brick edges; b) plot profile of the line ROI showing a brick length of 160 pixels (corresponding to a frequency of 0.006/pix); c) Fourier spectrum of the plot profile, the strong peak at 0.125/pix corresponds to a length of 8 pixels, about the width of the cement between the bricks.

or adapted for a very different area, and potentially lead to new insights there that would not have been gained without this new way of seeing old data.

Chapter 8

Conclusions

The interface where biology and physics meet is arguably one of the most exciting areas of science today. Biophysics takes many forms, from the study of the group behaviour of animals with models from fluid dynamics, to the modelling of body structures with tensegrity, to the study of the physical properties of cells and many more examples - the use of methods and concepts from physics has the potential to shed new light on old questions in biology. One of these areas is automated image analysis, the use of computers to extract data from images using specially developed algorithms.

In this thesis we develop methods for analysing microscopy images of collagen (a ubiquitous and important structural macromolecule with fibrous organisation and a characteristic banding pattern found in all multicellular animals) in a controlled and reproducible way and with minimal user input. The freely available software package Image SXM is used for this, and code is written in the inbuilt macro language based on the programming language Pascal. The focus of this thesis is the development of analysis strategies. No experimental results are presented.

The microscopy techniques by which the images were obtained were Atomic Force Microscopy (AFM), Polarised Light Microscopy (PLM), Second Harmonic Generation (SHG) and Transmission Electron Microscopy (TEM).

Images are ubiquitous in science and in most other areas of life, and vision more than any other sense is important in understanding the world. Seeing an image makes it possible to take in a wealth of information that could not easily be contained in words or numbers. Historically, new imaging techniques have usually led to leaps of new understanding. Hence, images are crucial as raw data in many diverse areas of science, from life sciences to earth sciences, and from astronomy to medicine. Yet, while looking at an image can reveal much that a description could not contain, there is more information in an image than can be seen by eye.

Our visual system has developed specifically to help us survive in the natural environment, and is very specialised in this task. It is excellent at recognising patterns and carries out complex calculations to make contextual sense of the visual stimuli, but is less good at being aware of the nature of the direct visual input, such as absolute colour values or brightness, degree of randomness, frequency elements, and so on. In visual analysis alone much information is lost. A computer, in contrast, deals with such image information much better than humans. Since images are valuable data, extracting all or most of the information in them gives a better return on the resources that were required to produce

them.

Image analysis by visual inspection is intrinsically prone to errors. Apart from the unavoidable bias towards seeing patterns (real or imagined), people are influenced by many factors, such as light conditions, seasons, mood, tiredness, hunger, health state, excitement or boredom, and this impacts their judgment. Reproducibility of decisions cannot be guaranteed from one hour to the next. This problem can be avoided by the use of computers. Furthermore, the volume of data produced is becoming increasingly large in all areas of science, and it is becoming less feasible for a large volume of images to be examined and evaluated by a human. For these reasons, automated computerised image analysis is a valuable, interdisciplinary, tool that is becoming more and more essential to evaluating image data.

Image analysis, though not new, is not a unified field. Many researchers attempt to develop their own solutions for very specific purposes, but there is no well-defined body of core knowledge, skills, methods or language, although limited common concepts do exist. In this work a range of concepts have been employed and analysis methods incorporating different principles have been developed.

The division of this thesis into chapters by microscopy technique is based on the different analysis strategies each technique demanded.

In the AFM chapter, an analysis strategy is developed that measures the thickness of a collagen fibre, as well as the frequency elements that reflect the banding frequency, and requires minimal user input.

The PLM chapter details the development of a different analysis strategy that attempts to capture more textural characteristics of the image field. This analysis introduces three metrics based on the autocorrelation function of parts of the image. No user input is needed for this analysis. The metrics are tested and shown to be able to discriminate successfully between visually different groups of images.

The SHG chapter illustrates the use of the analysis strategy developed in the PLM chapter on images of the same samples taken by a different microscopy technique, and demonstrates the robustness of the analysis. The results both reflect the differences in the microscopy technique and demonstrate that the ability to discriminate between the image groups is retained. This suggests a wider applicability of the analysis strategy to other structurally similar image types.

The TEM chapter exemplifies a case where the attempts to analyse the images was not successful. The images are investigated in detail and found to contain the triple problems of low contrast, high noise and broken outlines, sufficient to render all analysis attempts fruitless. A method of quantitatively assessing the relative success of analysis attempts, as well as image quality, is introduced and discussed.

In a wider context, it is hoped that the thoughts and ideas developed in this thesis may contribute to a greater awareness of the possibilities of image analysis, and provide a small impulse for a wider concerted effort to develop a common language and a body of knowledge that can be easily taught and understood across the disciplines.

It is also hoped that the specific analysis strategies developed here will be taken up and

used in the analysis of larger volumes of real data, allowing more in-depth observation of the behaviour of the metrics (in the case of the PLM and SHG analysis) and testing in earnest the analysis' ability to categorise images by their textural properties. While the analysis methods can deal with images of different content, as long as they show structural similarity, it may be interesting to develop the strategies presented here further, and to adapt them to different image types.

The work in this thesis shows that image analysis techniques can be powerful tools in the evaluation of images obtained with a range of techniques, can be aimed at different image characteristics, and avoids the biases inherent in visual image evaluation.

Appendix A

Brief Summary of Some Image Analysis tools

A.1 Thresholding and Density Slicing

Thresholding, also referred to as binarisation, is a way to turn an image from grey scale into a black and white image. This is done by selecting a grey value as a cut-off point, a "threshold value", and setting every pixel with a grey value below this threshold to white, and every pixel with a grey value above the threshold to black. Judicious selection of the threshold allows the selective removal of elements such as artifacts or noise. As gradients and other background variations are common, some preprocessing may be necessary to remove background variation that might otherwise mask the features.
It is a limitation that straightforward thresholding can only segment the image into two parts, and keep or discard only elements at the bottom or the top of the grey value range. An improvement on thresholding is Density Slicing. In this procedure, a band of grey values can be selected, with a specified upper and lower threshold. Subsequently, there are several options: either to replace the value of the selected pixels with a chosen foreground grey value, and then either to leave the remaining pixels unchanged or to replace their value with a chosen background colour, or to leave the selected pixels unchanged, and then again either to leave the remaining pixels unchanged or to replace their value with a chosen background colour. This is a more powerful approach that also, for instance, allows the removal or selection of features that share similar grey values, but are not at either end of the histogram.
For illustration, see fig. A.1.

A.2 Erosion and Dilation

Erosion and dilation are operations carried out only on binary images. In erosion and dilation pixels get turned "off" (here from black to white) and "on" (here from white to black), respectively, depending on its adjacent 8 pixels. The number of black pixels around a central pixel is termed the "count" and ranges from 1 to 8. This condition has to be decided upon by the user.
In erosion a black pixel becomes white if at least the number of "count" white pixels surround it, for instance if a black pixel is surrounded by five white pixels and the count

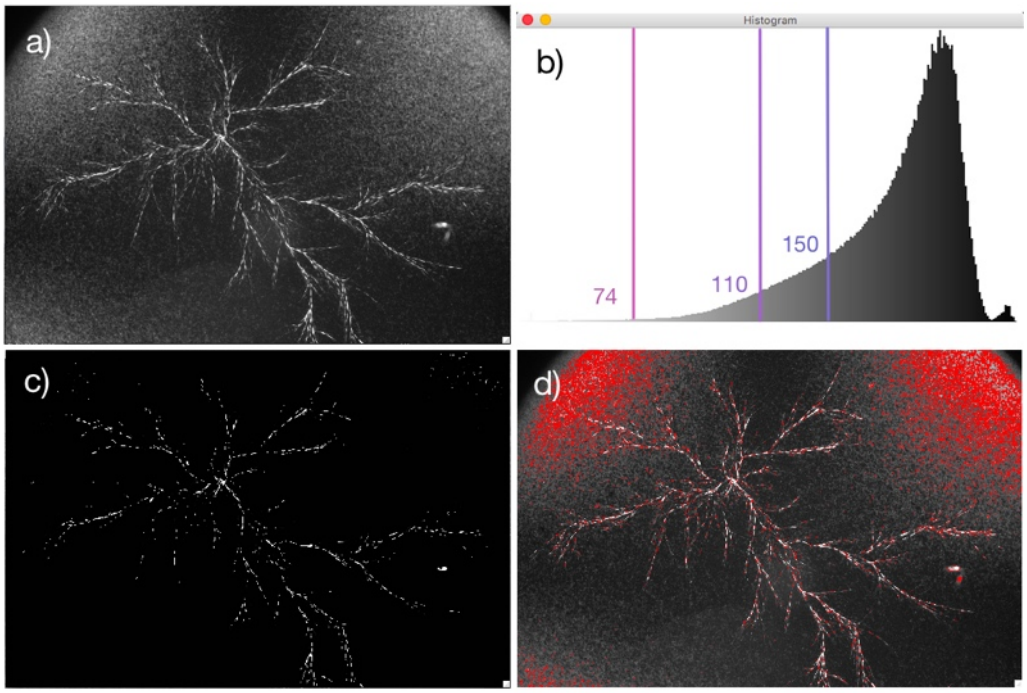


Figure A.1: Examples of thresholding and density slicing. a) original image (artifact, presumed to be houseplant pollen); b) histogram of a); c) image a) thresholded at grey value 74. All pixels with a higher grey value have been set to black, all pixels below 74 to white; d) image a) density sliced between grey values 110 and 150. The values between 110 and 150 are shown in red. These pixels can either be given a set grey value, leaving the remaining pixels unchanged, or can be left unchanged and the remaining pixels can be replaced by a given grey value.

is set to 3, then this pixel will be turned to white. But if for the same pixel the count was set to 6, then the pixel would remain black. Likewise, if the count is set to 3, a black pixel that is surrounded by two white pixels will not be turned white, but a black pixel surrounded by five white pixels will be. This is illustrated in fig. A.2.

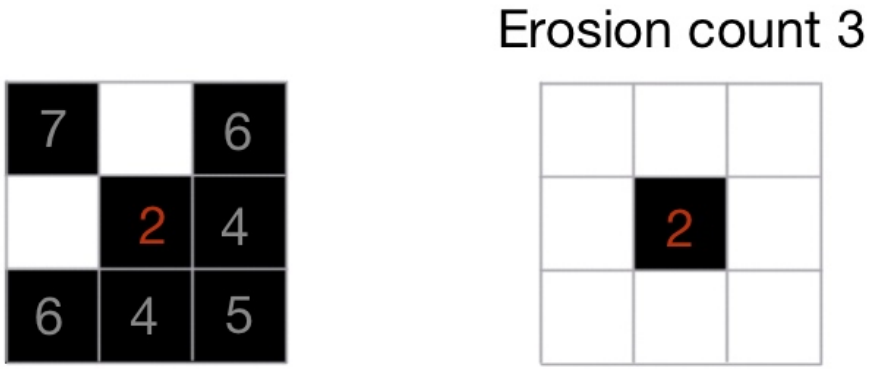


Figure A.2: Example of binary erosion. Only black pixels are considered in this operation. The numbers in the black pixels reqresent the number of white pixels touching. When the count is set to 3, a black pixel is turned white only if the number of white pixels touching it is at least 3. Therefore, the only pixel that remains black is the central one, whose neighbourhood only contained two white pixels.

In dilation a white pixel gets turned black if at least "count" black pixels surround it. For example, if a white pixel is surrounded by six black pixels and the count is set to 4, then this pixel will be turned to black. If for the same white pixel surrounded by six black ones the count was set to 7, then the white pixel would remain white. With a count set to 3, a white pixel surrounded by two blacks will not be turned black, while a white pixel surrounded by five black pixels would. An illustration of this is shown in fig. A.3.

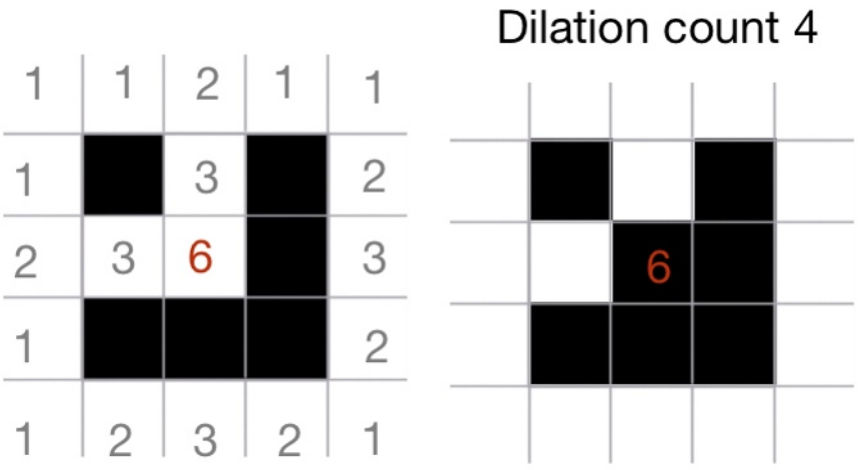


Figure A.3: Example of binary dilation. Only white pixels are considered in this operation. The numbers in the white pixels represent the number of black pixels in the immediate neighbourhood. When the count is set to 4, white pixels with at least 4 black pixels in the neighbourhood are turned black. In this case, only the central pixel whose neighbourhood contained 6 black pixels is turned black.

Erosion is useful in removing fuzzy edges or thin bridges between black areas, which then allows them to be identified as separate objects. The discretion of the operator in judging the amount of erosion is necessary, but also a potential source of inconsistency.
Dilation can fill in fuzzy edges and bridge gaps between close but incontiguous black areas. As with erosion, this can be useful, but is also subject to operator's decisionmaking.
Erosion and dilation are not the reverse of one another. Thus, an erosion followed by a dilation, or vice versa, does not return to the starting image. An improved image can sometimes be arrived at by a well-judged sequence of erosion and dilation steps in various combinations, but such complicated sequences are often only applicable to one specific image.

A.3 Kernel Filters and Convolution

In image processing, convolution is a process by which each pixel in the image is changed according to its neighbourhood through a kind of mask called a kernel. The kernel is a (usually small and square) matrix of integers. These integers are weighting factors for the neighbourhood pixels. The kernel is moved across the image pixel by pixel. The new grey value of the central pixel is determined by combining the weighted grey values of all the neighbourhood pixels together, and these result in a new and modified image. Fig. A.4 illustrates this and shows an example from a larger image. Convolution or kernel filters can be designed to modify images in a number of ways. Depending on the nature of the kernel, it is possible to detect edges, and smooth or sharpen an image to various degrees.

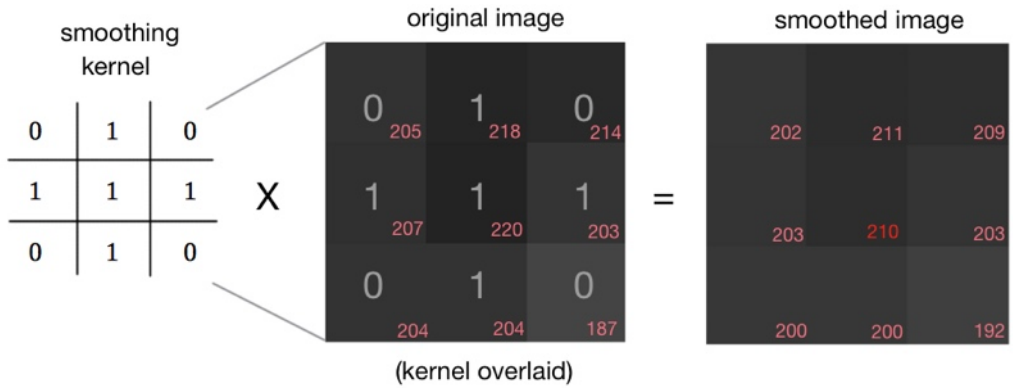


Figure A.4: Example of the convolution of an image with a kernel (in this case a simple smoothing kernel). The grey values of the image pixels are indicated in pink. When the kernel filter is applied to this image (which is part of a larger image), the central pixel's new grey value becomes $(1x218 + 1x207 + 1x220 + 1x203 + 1x204)/5 = 210$. The grey values in the other pixels are also changed, since the larger source image was smoothed as a whole.

A.4 Sobel filter

The Sobel filter is an edge detection filter. It is based on two 3x3 kernels (fig. A.5) that are convolved with the image. (For the explanation of convolution see A.3.) One

kernel detects vertical gradients by applying weighting factors that emphasise gradients between the left and right side of the central pixel, the other detects horizontal gradients by emphasising the difference between the pixels above and below the central pixel. This generates data for two separate images that are not usually shown. Both these image data are then combined to generate one image of edges, independent of their direction. Fig. A.6 shows an example of this.

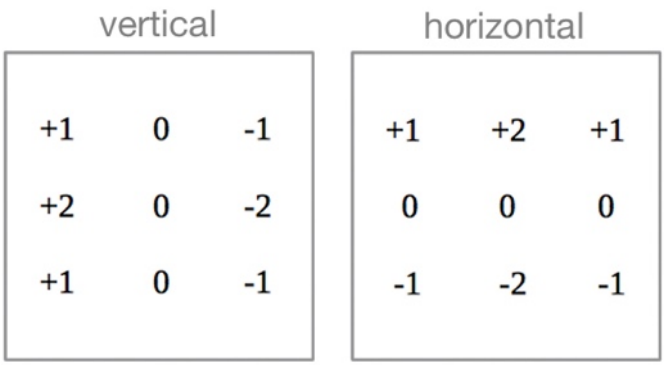


Figure A.5: Kernels of the Sobel filter: the kernel emphasising vertical edges is on the left, the kernel emphasising horizontal edges on the right.

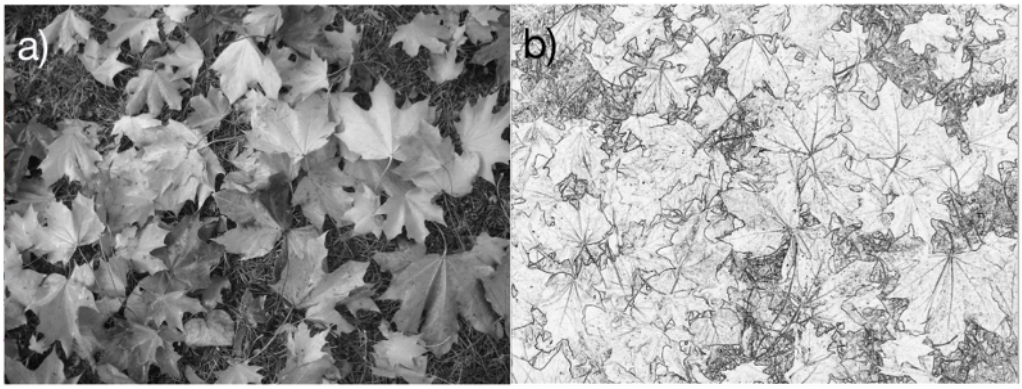


Figure A.6: a) Original image containing many steep grey value gradients ; b) after application of the Sobel filter: only the edges remain.

A.5 Watershed

The Watershed ([141]) is an effective method for separating objects that touch or have a partial overlap, if the overlap is not too extensive and the objects are mostly convex (see fig. A.7). The method first requires a binary image and then calculates the shortest distance to the edge of the boundary for each pixel inside the blob. It assigns a grey value to each pixel that increases with the distance from the edge. The result is the distance map in fig. A.7 b). The local maxima of the distance map are called the ultimate points, as they can also be arrived at by ultimate erosion from the edges (c) in A.7). The ultimate points then get progressively dilated until either the blob boundary or the growing region

around another ultimate point is reached. The collision line between two growing ultimate point regions is the watershed line.

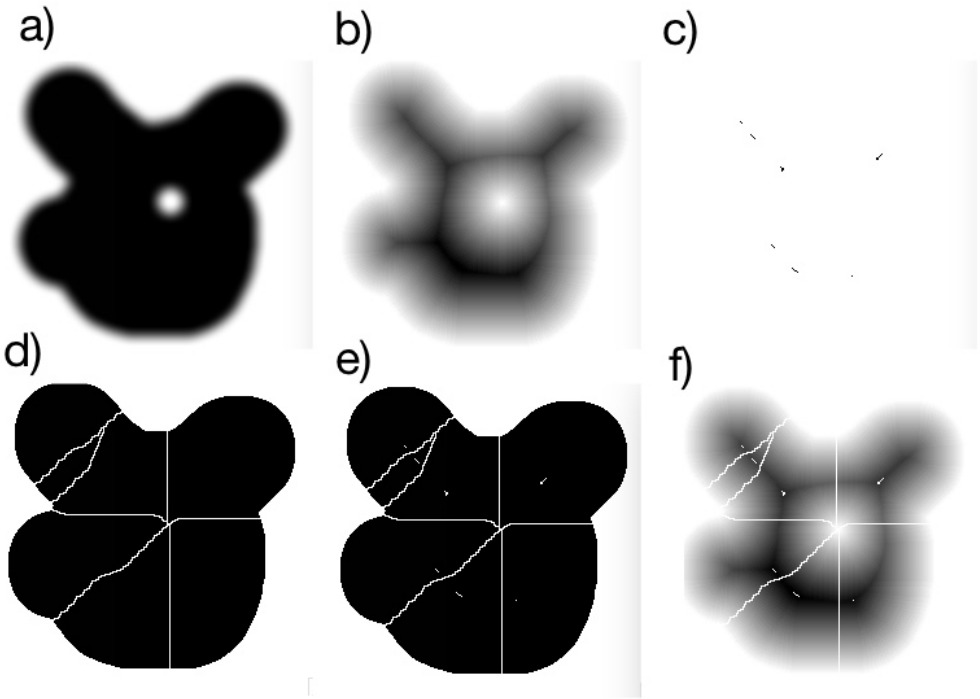


Figure A.7: Demonstration of the watershed segmentation method. An irregular blob is shown in a). For the distance map, b), the pixels inside the blob are assigned grey values as a function of their nearest distance to a boundary. The ultimate points in c) are derived from sequential erosion of the blob edges and are the local maxima of b). The watershed lines are shown in d), for clarity in e) the watershed lines are shown together with the ultimate points, and in f) the distance map, the ultimate points and the watershed lines are overlaid to show their relationship.

A.6 Particle Analysis

The Particle Analysis is a legacy from NIH Image and an inbuilt function of Image SXM. As the name suggests, its function is to analyse "particles", that is black objects in a binary image. The "Measurement Options" dialogue box (fig. A.8) offers the user a choice of measurements for the analysis. Not all of these are relevant to the Particle Analysis, as these options also pertain to the inbuilt "Measure" function, which also deals with grey value images.

Further options for the Particle Analysis specifically are given in the "Particle Analysis options" dialogue box (fig. A.9), such as the options to label or outline particles in the results image, to exclude particles touching the image edges, or to include interior holes in the count.

The numerical results can be displayed in a text window and saved or exported. Internally they are stored in dedicated arrays that can be accessed from the macro language. Fig. A.10 shows an example of an image after the Particle Analysis, together with the

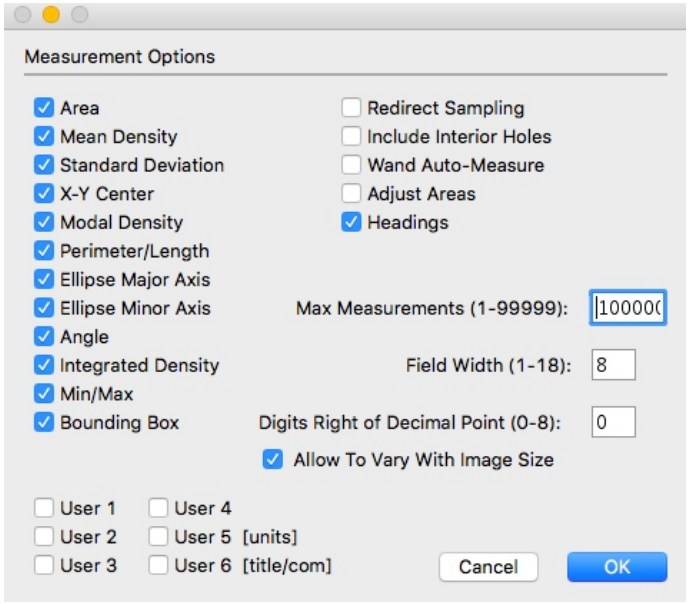


Figure A.8: The Measurement Options dialogue box of the Particle Analysis and Measure functions.

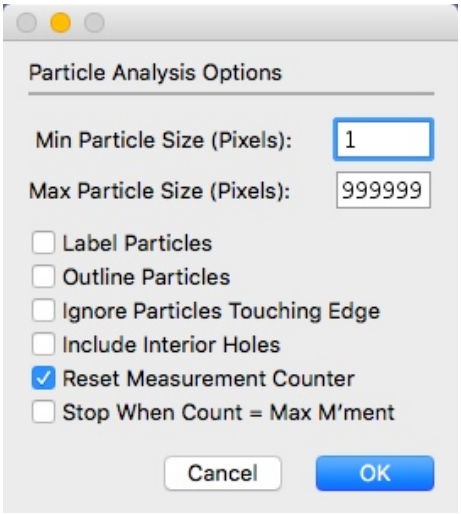


Figure A.9: The Particle Analysis Options dialogue box.

measurements taken from the image.

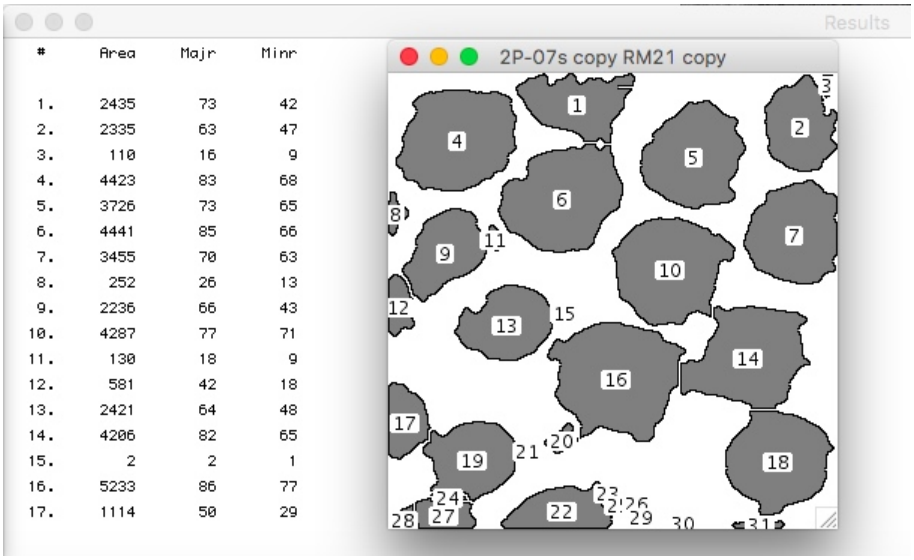


Figure A.10: Example of an image after the Particle Analysis has been run, and the Results window with the measurements from that image.

A.7 Rotating Mask

The Rotating Mask (RM) is a smoothing filter with edge preservation. In general terms, for every pixel a neighbourhood of a certain size is divided into two equal parts, and each is examined for its inhomogeneity with respect to a certain criterion (for example variation in grey value). In most cases, one part is more inhomogeneous than its complement. The part in which the inhomogeneity is most pronounced is selected. Its counterpart is then used to assign a new grey value to the central pixel, for example the average grey value of the complement part. The underlying assumption is that the highest variability in a subset will always occur across a boundary. If the pixels near the boundary get assigned a grey value of the opposite side, a separation of grey values will occur at the boundary, and the edge will not be blurred. The principle of this operation is illustrated in fig. A.11. The RM used here puts a ROI around each pixel and divides the ROI into two by a rotating diagonal. In each half (subset) the standard deviation of grey values is measured and stored. When the diagonal has completed a 180 degree rotation, the complement of the subset with the highest variability in grey values is selected and its mean grey value is used as the new grey value for the pixel. Fig. A.12 shows a flow diagram of our implementation of the Rotating Mask.
Fig. A.13 shows an example of a Rotating Mask being applied. The original image in a) is grainy and a definitive edge between blobs and background cannot be seen anywhere. After the Rotating Mask has been applied in b), some crisp edges have appeared.

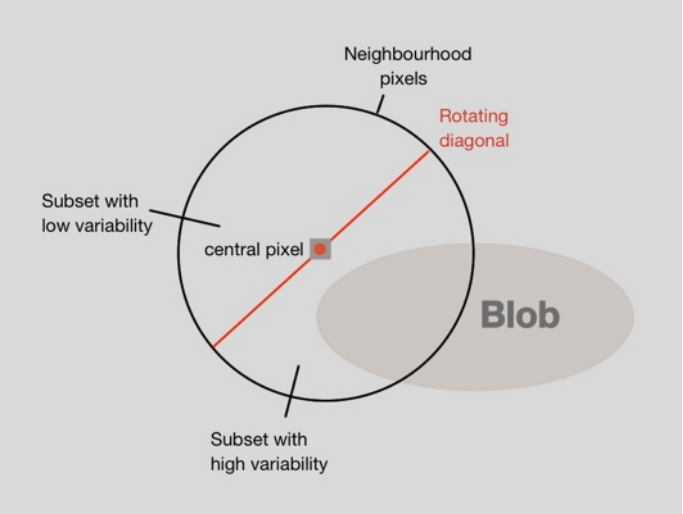


Figure A.11: Principle of a Rotating Mask. The environment of a pixel is divided into two parts, each of which is examined with respect to its homogeneity. As the subset of pixels covers a blob, the inhomogeneity increases. The subset that covers the blob to the greatest extent is likely to have the largest degree of inhomogeneity and its counterpart is likely most tangential to the blob. Thus, the new pixel grey value is assigned in relation to this counterpart, for example the average grey value.

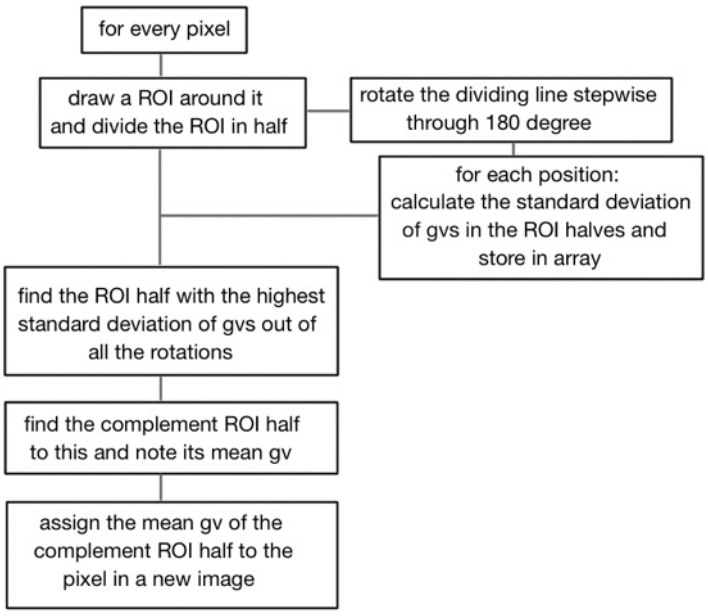


Figure A.12: Flow diagram of the Rotating Mask implementation in the TEM analysis. Standard deviation of grey values is used as inhomogeneity criterion, and the mean grey value of the complement of the ROI half with highest inhomogeneity is used for the pixel value in the filtered image.

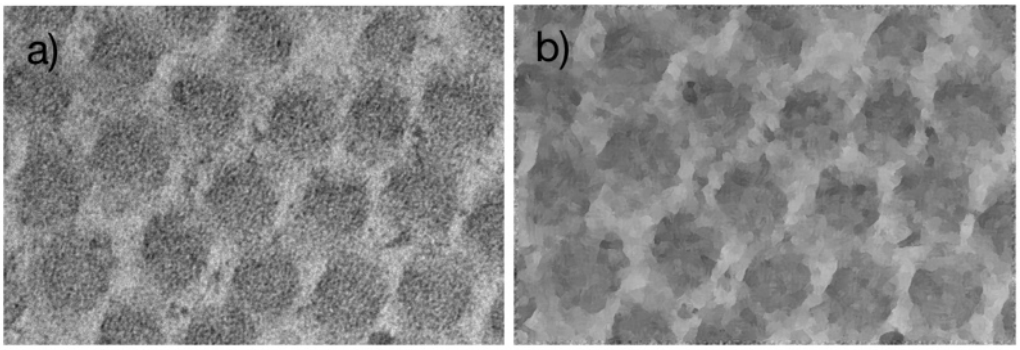


Figure A.13: Example of the effect of the Rotating Mask. a) original image with no sharp edges; b) after the Rotating Mask has been applied, some edge features can be discerned.

A.8 Hough Transform

The classical Hough transform ([142]) is a method of detecting features of a particular, known form in an image. This is done by a procedure which uses a multidimensional space (the Hough Space), in which each dimension consists of a parameter necessary to describe the feature that should be detected. Each coordinate point in the image is assumed to be part of the known feature, and the parameters of all features that might contain the coordinate point are calculated and entered into the Hough Space. The Hough Space is divided into discrete "cells", which are filled up with the calculated parameters. As this is done for all coordinate points, other points that are part of the same feature also have the same possible parameters associated with them. In this case, the "cells" of the Hough Space for these specific parameters get filled up also by the calculations for those other points. In this way, the Hough Space is filled up and acquires "hot points" (accumulators), where the shapes in this parameter space overlap. This cumulative filling up of specific areas in the Hough Space is also called a voting procedure, as each hit of a cell can be seen as a vote for that set of parameters. Fig. A.14 shows the parameters that can be set for the Hough Transform within IMAGE SXM. A dummy image with roundish shapes illustrates the results of the Hough Transform (fig. A.15).

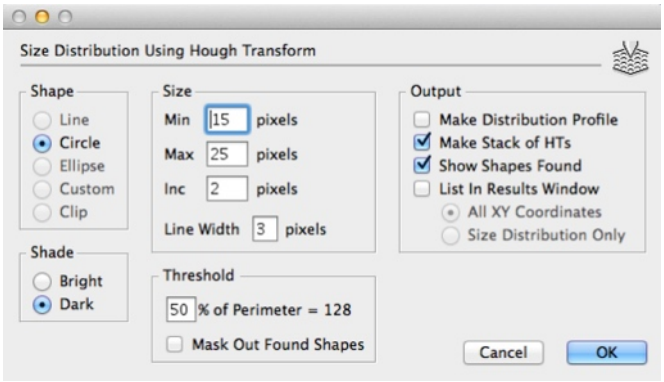


Figure A.14: User interface of special HT analysis that detects size distribution in IMAGE SXM. Sizes from 15 to 25 pixels were chosen.

The HT is not sensitive to noise or breaks in the features, as the decision where in Hough

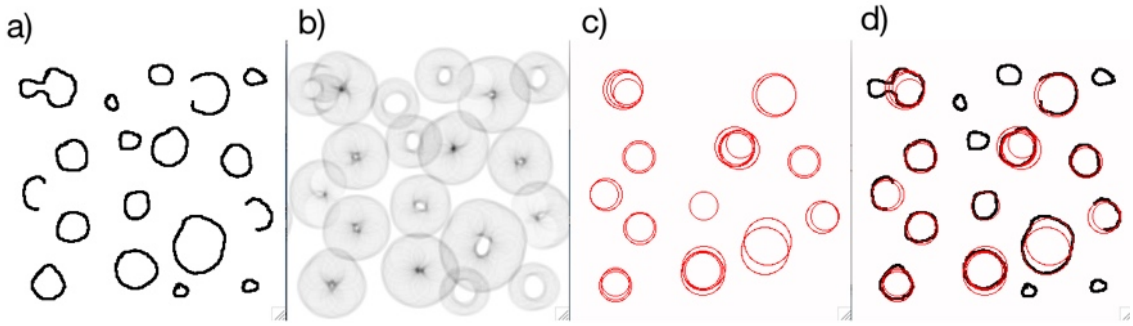


Figure A.15: Illustration of the Hough Transform: a) dummy image with arbitrary black shapes on a white background, on average 40 pixels in diameter; b) Hough Transform result for 19 pixel diameter circles (one of a stack from 15 to 25 pixels diameter, not shown), note that a central dark accumulator forms for objects that are close to the size of the applied diameter, while for objects of a different size the accumulator does not form; c) cumulative result: circles drawn round the accumulator points with the diameter at which they were found; d) overlay image of original with the result of the search for circles from 15 to 25 pixels in size.

Space a cut-off should be made (i.e. how many "votes" a cell needs to have before it is regarded as part of a real feature) is up to the user. However, it is not possible to distinguish between different reasons for a less-than-perfect match. Thus, a noisy or low contrast image with perfect, unbroken features presents the same difficulty for the HT as a crisp, high contrast image with very broken features.
In order to simplify the image features to more closely resemble geometric shapes, some preprocessing steps, such as edge detection or thresholding, are often advisable.
In the classical HT the shapes are described in parametric form, which is feasible for features described by few parameters, such as lines, simple curves and ellipses. In the generalised HT ([143])) features that cannot be parametrically described can be detected as well, but in this case a look-up table is used, which is created by the relative position of points on a prototype feature to a reference point within it.

A.9 Fast Fourier Transform - FFT

Every continuous, unambiguous function can be transformed into a series of sin and cos functions with distinct frequency, phase and amplitude. An image can be regarded as a two-dimensional continuous, unambiguous function, where the grey values in each pixel (0-255 in 8 bit images) represent the dependent variable, and the x and y coordinates of the pixel the independent variables.
The frequency and phase elements of the series of sin and cos expressions that describe the original curve (in one dimension) or image (in two dimensions) can be separated and displayed individually. The "power spectrum" displays the power at each frequency, i.e. how much of it there is at each frequency in the curve or image. In two dimensions, the power image shows which frequencies are in the image (relative to image size) and also their orientation (the orientation in the power image is perpendicular to the orientation in the image), and the grey values represent the corresponding amplitudes. The phase

image is usually not shown, since it bears little information accessible to observation, and the power image is often for simplicity referred to as the FFT image. The centre position in the power image is the lowest frequency (the largest object, the whole image), with increasing frequencies (smaller objects) towards the outside.
Both the power and phase parts can be recombined in an "inverse Fourier Transform" (iFFT) to reconstruct the original image. However, it is possible to manipulate the power image and thereby reconstruct the original image with some alterations. This potential has many uses. Among the most common are high, low and band pass filters, which selectively remove either large image features, or small scale image elements (such as noise), or image elements that fall within a specific frequency band, respectively. The filtering is achieved by selecting areas in the FFT image to be included or excluded in the reconstruction of the iFFT, usually by drawing into the FFT image. Fig. A.16 shows examples of these filters.
If image elements are characterised by distinct frequencies, they produce recognisable features in the FFT image. These features, appearing usually as lines or dots, can be specifically manipulated and thereby taken out of an image, or the rest of the image taken away to isolate them. This is a very powerful way of manipulating images that cannot be achieved with any other method. Figs. A.17 and A.18 illustrate this process. Fig. A.17 shows the unaltered FFT of the original image, followed by various sketched in masks. Fig. A.18 shows first the original image, followed by the iFFTs of the various masks in fig. A.17. (Only the outlines of the masks have been drawn for illustration purposes. In the application the outlines would be filled. Note that a black mask indicates to the software that only the FFT areas underneath the mask should be used for the iFFT, and a white mask indicates that all areas except the area underneath the mask should be used. This is equivalent to painting the area around the location of the white mask black.)
FFT-based image manipulation techniques are manifold and uniquely powerful tools for dealing with a variety of image data, and are therefore very widely employed.

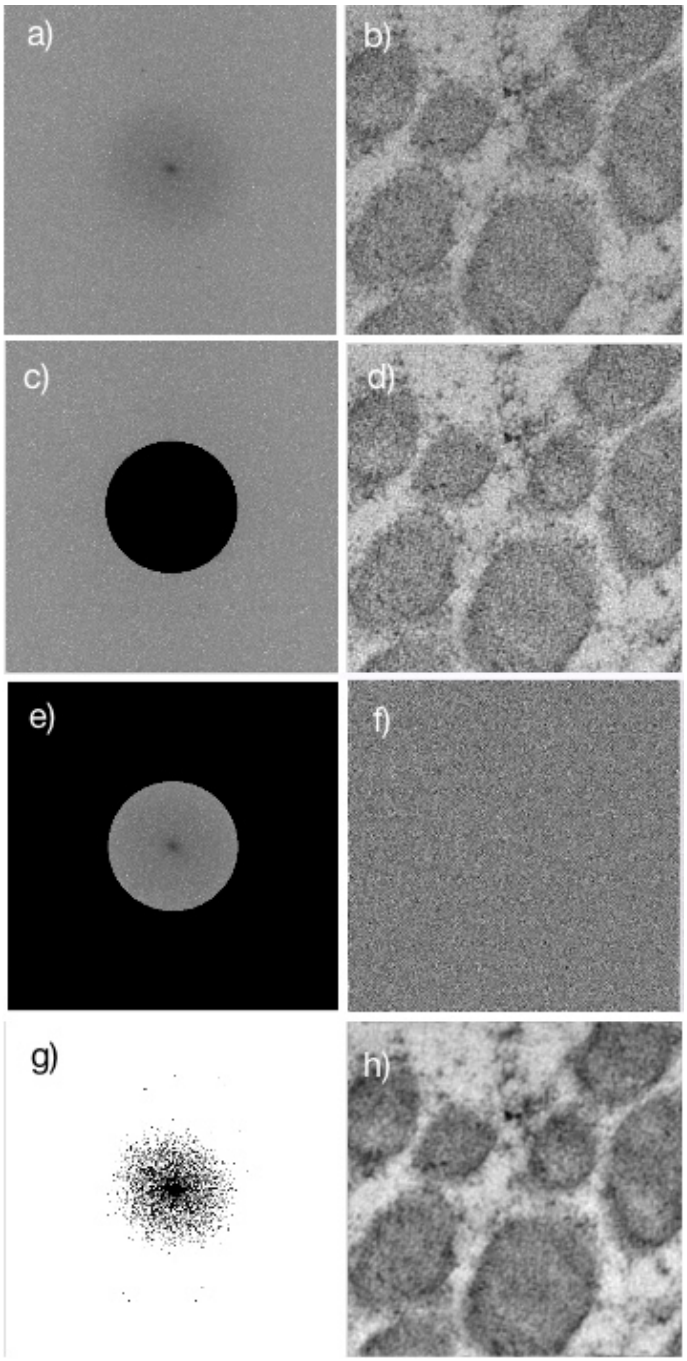


Figure A.16: Various filters based on manipulation of the FFT image. a) unmodified FFT image of b), low frequencies are in the centre, high frequencies toward the periphery; b) the original image; c) low pass filter: a black circle is drawn around the centre of the FFT, this acts as a mask that tells the software only to use the data underneath the black area for the iFFT; d) iFFT of c), only low frequencies (large objects) remain in the image; e) high pass filter: the black mask covers the entire FFT image except the area of the low freqencies around the centre; f) iFFT of e), only small scale structure is left in the image; g) filter based on thresholding the FFT image: this allows frequencies to be selected that occur above a certain power, independent of their length scale; h) iFFT of g)

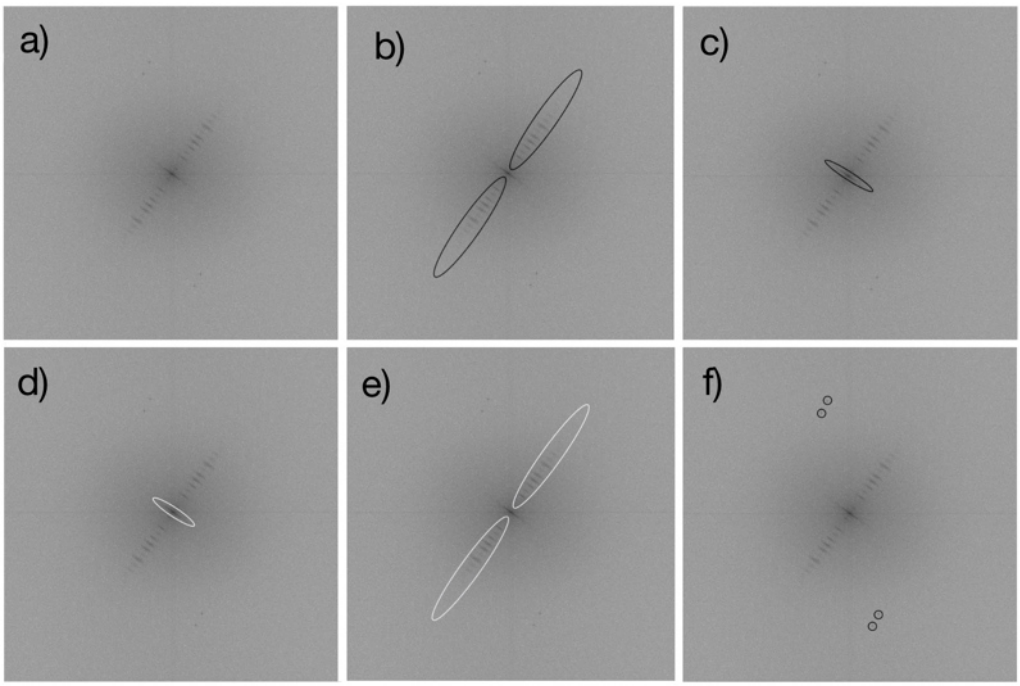


Figure A.17: Examples of masks used to separate image features by frequency. a) unaltered FFT image of the original (a) in A.18); b) mask selecting the "ladder" at around 45 degrees; c) mask selecting the unbroken diagonal through the centre; d) mask excluding the diagonal through the centre; e) mask excluding the "ladder" at 45 degrees; f) mask selecting the set of two double points. (Note that the masks are drawn in outline only, to allow to see what would be underneath.

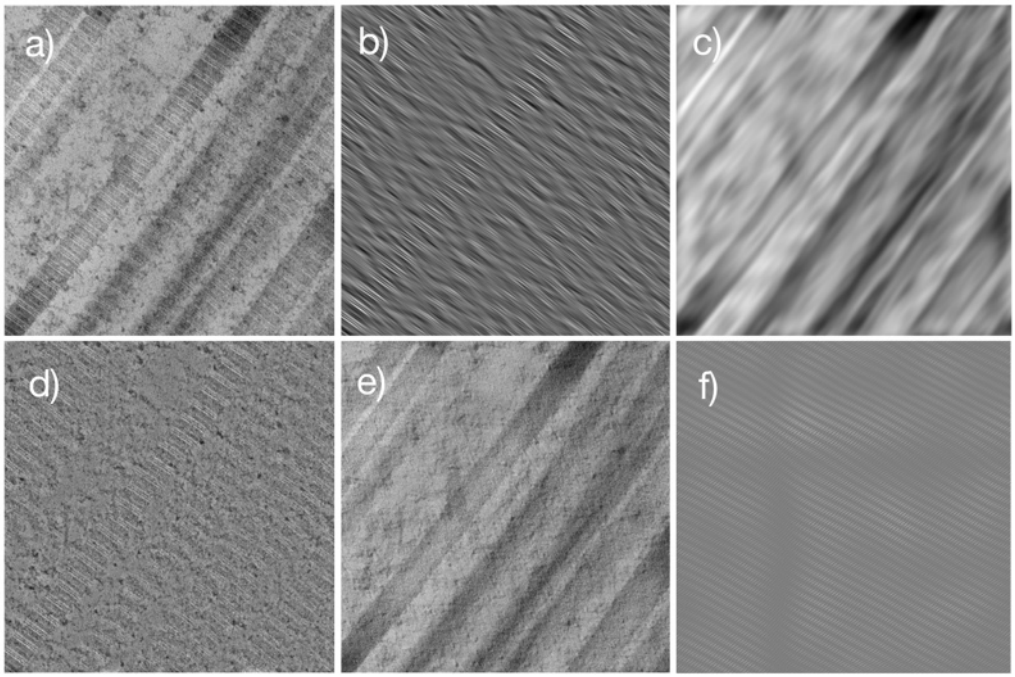


Figure A.18: Images corresponding to the FFT images and masks in fig. A.17. a) original image, unaltered; b) compared to a), this image has lost all but the banding pattern in the collage fibres. hence the 45 degree "ladder" in the image above originates from the collagen banding; c) in comparison to a), here all features have been lost, except the overall shapes of the long fibres, hence the diagonal through the centre represents the large-scale fibre shapes; d) the overall fibre shapes have been removed from this image, including the light and dark areas at large scales, leaving mainly the banding in a uniform background; e) the fibre banding is the only thing that has been removed from this image, leaving the long fibre shapes and the rest of the image features intact; f) the double points in fig. A.17 are sharp and clearly defined, pointing at very exact frequencies. Isolating these reveals a wave throughout the image near 50 Hz, commensurate with electrical noise.

Appendix B

Macro Codes

B.1 AFM Macro

```
macro 'AFM measurements';

{
This macro file lets the user draw a line ROI along the middle of a collagen
     fibre. The macro measures the periodicity of the fibre by means of an
    FFT and saves the FFT plot data in a file called 'filename-PV' in the
    same location as the source image.

The macro also attempts to find the fibre edges to the left and right of the
     user-drawn ROI in three locations along the ROI: the centre, the
    starting point and the end point.

It creates a new copy of the image window, into which it draws a line to
    represent the user ROI and also green dots where it locates the fibre
    boundaries. Red dots along the image margins are control points that
    indicate where secondary ROIs have been drawn along which the fibre edges
     were searched.

Additional new windows are:
1. a text window showing individual measurements of fibre half-diameters,
    diameters and average, as well as standard deviation,
2. an image window that shows the plot data line from which the FFT is
    calculated
3. the plot data graph of this line,
4. the FFT graph from this
}


{Global variables:}

var

xorig, yorig, xend, yend : integer;      {The start and end points of the
    secondary ROI}
halfdiam, diam1a, diam1b, diam2a, diam2b, diam3a, diam3b : real; {half the
    fibre diameters}

procedure drawreddot;

{draws a red dot - gets called from within drawROIquad1, 2, 3 and 4 -
    indicates where the ROIs drawn there meet the edge of the image}

begin

      setpalette('Grayscale',1);       {this puts red at a gv of 1}
      Makeovalroi(xend-4,yend-4,9,9);
      setforegroundcolor(1);
```

```
        fill;
        setforegroundcolor(255);
        killroi;

end;

{————————————————————————————}

procedure drawgreendot;

{draws a green dot - also gets called from within drawROIquad1, 2, 3 and 4 -
     indicates where the findstep procedure finds the fibre edges}

begin

        setpalette('Grayscale',2);        {this puts green at the gv of 2}
        Makeovalroi(xdot-2,ydot-2,5,5);
        setforegroundcolor(2);
        fill;
        setforegroundcolor(255);
        killroi;

end;

{————————————————————}

procedure doyinvert;

{Y invert has to be switched off for the angles to work (the ISXM default is
     switched on) - makes a ROI in a little window and checks the coordinates
     - depending on their values it inverts the y axis or not}

var

        coordnum : integer;
        origtop : boolean;
        picidy : integer;

begin

        setnewsize(20,20);
        makenewwindow('invert');
        picidy := pidnumber;
        choosepic(picidy);
        makeroi(0,0,10,10);
        coordnum := ncoordinates;   {ncoordinates fills the xcoordinates and
    ycoordinates inbuilt arrays, coordnum is the number of coordinates}
        killroi;
        origtop := (ycoordinates[1] = 0);   {this means "if ycoordinates[1]
    equals 0, then origtop is "true"}
        close;
        inverty(origtop);      {inverts Y if origtop is true}

end;

{————————————————————————————}

procedure findstep;

{secondary ROI has just been drawn, and this procedure tries to find the
    first steep peak in the plot data by comparing the gv of pixels '
    steplength' apart and decides there is an edge if the difference is
    greater than 'stepthresh'
```

```
get called from within drawROIquad procedures}

var

count, ppv, min, max : integer;
stepthresh, steplength : integer;
delta1, delta2 : real;
stoploop : integer;
ROIlength : integer;
k : integer;

begin

      resetcounter;

      getplotdata(count,ppv,min,max);

      ROIlength := rLength[1];

      for k := 1 to ROIlength do begin

            ruser1[k] := plotdata[k];

      end;

      stepthresh := 5;
      steplength := 5;

      k := steplength + 1;            {so that the first step starts
    steplength away from the edge}
            repeat
                  delta1 := ruser1[k] - ruser1[k - steplength];
                  delta2 := ruser1[k] - ruser1[k + steplength];
            k := k + 1;
            until (delta1 > stepthresh) and (delta2 > stepthresh);

            halfdiam := k;        {necessary to fill the text window with a
    sequence of different halfdiams}

end;

{-------------------------------------}


procedure drawROIquad1;

{draws a secondary ROI into the 1. quadrant, which is perpendicular to the
    user ROI drawn into the 4th quadrant
from in there it draws a red dot where the ROI meets the edge, calls the
    procedure that finds the edge and draws a green dot where the fibre edge
    is believed to be}

var

h1, h2 : real;
delta : real;              {the angle of the 2nd ROI used for drawing green
    dots}
xdot, ydot : real;

begin

      choosepic(picid2);
```

```
      h1 := abs(yorig / cos(alpha));
      h2 := abs((picwidth - xorig) / cos(beta));

      if h1 < h2 then begin      {Hypotenuse below corner, r cos(alpha) = x0,
     calc y}
            xend := xorig + h1 * abs(sin(alpha));
            yend := 0;
      end else
       if h1 > h2 then begin       {Hypotenuse above corner, r cos(beta) = y0
    , calc x}
            xend := picwidth;
            yend := yorig - h2 * abs(sin(beta));
      end else
      if h1 = h2 then begin      {Hypotenuse goes through corner}
            xend := picwidth;
            yend := 0;
      end;

      drawreddot;

      setlinewidth(5);

      makelineROI(xorig,yorig,xend,yend);

      measure;
      delta := rAngle[1];        {needed for the green dot}
      resetcounter;

      findstep;

      killROI;

      setlinewidth(1);

      delta := delta * (3.14159265/180);       {convert from degree into
    radians}
      xdot := xorig + (halfdiam * cos(delta));
      ydot := yorig - (halfdiam * sin(delta));

      drawgreendot;

end;

{————————————————}

procedure drawROIquad2;

{draws secondary ROIs into the 2. quadrant, which is perpendicular to the
    user ROI drawn into the 1st quadrant
from in there it draws a red dot where the ROI meets the edge, calls the
    procedure that finds the edge and draws a green dot where the fibre edge
    is believed to be}

var

h1, h2 : real;
delta : real;              {the angle of the 2nd ROI used for drawing green
    dots}
xdot, ydot : real;

begin
```

```
      choosepic(picid2);

      h1 := abs(yorig / cos(alpha));
      h2 := abs(xorig / cos(beta));

      if h1 < h2 then begin      {Hypotenuse below corner, r cos(alpha) = x0,
    calc y}
            xend := xorig - h1 * abs(sin(alpha));
            yend := 0;
      end else if h1 > h2 then begin      {Hypotenuse above corner, r cos(
   beta) = y0, calc x}
            xend := 0;
            yend := yorig - h2 * abs(sin(beta));
      end else if h1 = h2 then begin      {Hypotenuse goes through corner}
            xend := 0;
            yend := 0;
      end;

      drawreddot;

      setlinewidth(5);

      makelineROI(xorig,yorig,xend,yend);

      measure;
      delta := rAngle[1];       {needed for the green dot}
      resetcounter;

      findstep;

      killROI;

      setlinewidth(1);

      delta := delta * (3.14159265/180);      {convert from degree into
   radians}
      xdot := xorig + (halfdiam * cos(delta));
      ydot := yorig - (halfdiam * sin(delta));

      drawgreendot;

end;

{-------------------}

procedure drawROIquad3;

{draws secondary ROIs into the 3. quadrant, which is perpendicular to the
    user ROI drawn into the 2nd quadrant
from in there it draws a red dot where the ROI meets the edge, calls the
    procedure that finds the edge and draws a green dot where the fibre edge
    is believed to be}

var

h1, h2 : real;
delta : real;             {the angle of the 2nd ROI used for drawing green
    dots}
xdot, ydot : real;

begin
```

```
      choosepic(picid2);

      h1 := abs((picheight - yorig) / cos(alpha));
      h2 := abs(xorig / cos(beta));

      if h1 < h2 then begin      {Hypotenuse below corner, r cos(alpha) = x0,
     calc y}
            xend := xorig - h1 * abs(sin(alpha));
            yend := picheight;
      end else if h1 > h2  then begin      {Hypotenuse above corner, r cos(
    beta) = y0, calc x}
            xend := 0;
            yend := yorig + h2 * abs(sin(beta));
      end else if h1 = h2 then begin       {Hypotenuse goes through corner}
            xend := 0;
            yend := picheight;
      end;

      drawreddot;

      setlinewidth(5);

      makelineROI(xorig,yorig,xend,yend);

      measure;
      delta := rAngle[1];        {needed for the green dot}
      resetcounter;

      findstep;

      killROI;

      setlinewidth(1);

      delta := delta * (3.14159265/180);       {convert from degree into
    radians}
      xdot := xorig + (halfdiam * cos(delta));
      ydot := yorig - (halfdiam * sin(delta));

      drawgreendot;

end;


{-------------------}

procedure drawROIquad4;

{draws secondary ROIs into the 4. quadrant, which is perpendicular to the
    user ROI drawn into the 3rd quadrant
from in there it draws a red dot where the ROI meets the edge, calls the
    procedure that finds the edge and draws a green dot where the fibre edge
    is believed to be}

var

h1, h2 : real;
delta : real;              {the angle of the 2nd ROI used for drawing green
    dots}
xdot, ydot : real;

begin
```

```
      choosepic(picid2);

      h1 := abs((picheight - yorig) / cos(alpha));
      h2 := abs((picwidth - xorig) / cos(beta));

      if h1 < h2 then begin      {Hypotenuse above corner}
            xend := xorig + h1 * abs(sin(alpha));
            yend := picheight;
      end else if h1 > h2  then begin      {Hypotenuse below corner}
            xend := picwidth;
            yend := yorig + h2 * abs(sin(beta));
      end else if h1 = h2 then begin      {Hypotenuse goes through corner}
            xend := picwidth;
            yend := picheight;
      end;

      drawreddot;

      setlinewidth(5);

      makelineROI(xorig,yorig,xend,yend);

      measure;
      delta := rAngle[1];        {needed for the green dot}
      resetcounter;

      findstep;

      killROI;

      setlinewidth(1);

        delta := delta * (3.14159265/180);       {convert from degree into
    radians}
        xdot := xorig + (halfdiam * cos(delta));
        ydot := yorig - (halfdiam * sin(delta));

        drawgreendot;

end;

{-----------------}

procedure getperiod;

{elongates the user ROI data to the next higher power of 2 and then does an
    FFT on these}

var

expo, expochk, minval : real;        {exponent - for checking if pixel number
     is a round power or 2}

begin

        expo := ln(linelength) / ln(2);        {2 to the power of expo =
    linelength, expo is the exponent of 2 that results in linelength}

        expo := trunc(expo);              {we want an integer power of 2}

        expochk := ln(linelength) / ln(2);         {in case linelength is
    already a power of 2 ...}
```

```
    if (expo <> expochk) then begin
          expo := expo + 1;      {expo is now a full integer, if expo is
    not, then we want to extend the line length to the next power of 2
    upwards, so expo becomes 1 more}
          linelength := exp(expo * ln(2));  {This is 2 to the power of the
     new expo! See Green}
    end;

    showmessage(linelength:4:0);

    difflength := linelength - linelength1;      {for filling up difflength
     with made up values}

                    {both linebuffer and plotdata start with 0, not 1}

    for i := 0 to linelength1 - 1 do begin
          rUser1[i] := plotdata[i];   {transferring values from plotdata
    to rUser1}
    end;

    for i := (linelength1) to linelength - 1 do begin

              rUser1[i] := rUser1[(2*linelength1) - i ];  {This fills
    the rest of the array with a mirror image of what came before.}

    end;

    for i := 0 to linelength - 1 do begin
          linebuffer[i] := rUser1[i];
    end;

    setnewsize(linelength,2);
    makenewwindow('LineROI');
    putrow(0,0,linelength);          {Draws the new values into a narrow
    image window, this is necessary for the FFT}

    makelineROI(0,0,linelength,0);

    fft('foreward');
    {I'm not sure in which array this sits, so that the max could be
    extracted easily, but I don't have access to it and it CAN be exported as
     a text file!}

    setexport('plot values');
    export(title '-PV');

end;


{-------------------------------------------------}

procedure drawmainroi;

{for reference, draws a line in the copy image in the place corresponding to
     where the user drew the original ROI}

begin

    setlinewidth(3);
    moveto(x1,y1);
    lineto(x2,y2);
    setlinewidth(1);
```

```
end;

{-------------------------------------------------------------}

procedure drawallROIs;

{This procedure decides which ROI-drawing procedure is called in sequence
    from centre points to start and end points, and fills the diameter values
     into the text window 'Fibre diameters'.}

begin

      if quadr = 1 then begin

            xorig := x0;
            yorig := y0;

            drawROIquad2;
            diam1a := halfdiam;
            selectwindow('Fibre diameters');
            writeln('Diam1a is ',diam1a);

            drawROIquad4;
            diam1b := halfdiam;
            selectwindow('Fibre diameters');
            writeln('Diam1b is ',diam1b);


            xorig := x1;
            yorig := y1;

            drawROIquad2;
            diam2a := halfdiam;
            selectwindow('Fibre diameters');
            writeln('Diam2a is ',diam2a);

            drawROIquad4;
            diam2b := halfdiam;
            selectwindow('Fibre diameters');
            writeln('Diam2b is ',diam2b);


            xorig := x2;
            yorig := y2;

            drawROIquad2;
            diam3a := halfdiam;
            selectwindow('Fibre diameters');
            writeln('Diam3a is ',diam3a);

            drawROIquad4;
            diam3b := halfdiam;
            selectwindow('Fibre diameters');
            writeln('Diam3b is ',diam3b);

      end;

      if quadr = 2 then begin

            xorig := x0;
            yorig := y0;
```

```
        drawROIquad3;
        diam1a := halfdiam;
        selectwindow('Fibre diameters');
        writeln('Diam1a is ',diam1a);

        drawROIquad1;
        diam1b := halfdiam;
        selectwindow('Fibre diameters');
        writeln('Diam1b is ',diam1b);


        xorig := x1;
        yorig := y1;

        drawROIquad3;
        diam2a := halfdiam;
        selectwindow('Fibre diameters');
        writeln('Diam2a is ',diam2a);

        drawROIquad1;
        diam2b := halfdiam;
        selectwindow('Fibre diameters');
        writeln('Diam2b is ',diam2b);


        xorig := x2;
        yorig := y2;

        drawROIquad3;
        diam3a := halfdiam;
        selectwindow('Fibre diameters');
        writeln('Diam3a is ',diam3a);

        drawROIquad1;
        diam3b := halfdiam;
        selectwindow('Fibre diameters');
        writeln('Diam3b is ',diam3b);

end;

if quadr = 3 then begin

        xorig := x0;
        yorig := y0;

        drawROIquad4;
        diam1a := halfdiam;
        selectwindow('Fibre diameters');
        writeln('Diam1a is ',diam1a);

        drawROIquad2;
        diam1b := halfdiam;
        selectwindow('Fibre diameters');
        writeln('Diam1b is ',diam1b);


        xorig := x1;
        yorig := y1;

        drawROIquad4;
        diam2a := halfdiam;
        selectwindow('Fibre diameters');
        writeln('Diam2a is ',diam2a);
```

```
        drawROIquad2;
        diam2b := halfdiam;
        selectwindow('Fibre diameters');
        writeln('Diam2b is ',diam2b);


        xorig := x2;
        yorig := y2;

        drawROIquad4;
        diam3a := halfdiam;
        selectwindow('Fibre diameters');
        writeln('Diam3a is ',diam3a);

        drawROIquad2;
        diam3b := halfdiam;
        selectwindow('Fibre diameters');
        writeln('Diam3b is ',diam3b);

end;

if quadr = 4 then begin

        xorig := x0;
        yorig := y0;

        drawROIquad1;
        diam1a := halfdiam;
        selectwindow('Fibre diameters');
        writeln('Diam1a is ',diam1a);

        drawROIquad3;
        diam1b := halfdiam;
        selectwindow('Fibre diameters');
        writeln('Diam1b is ',diam1b);


        xorig := x1;
        yorig := y1;

        drawROIquad1;
        diam2a := halfdiam;
        selectwindow('Fibre diameters');
        writeln('Diam2a is ',diam2a);

        drawROIquad3;
        diam2b := halfdiam;
        selectwindow('Fibre diameters');
        writeln('Diam2b is ',diam2b);


        xorig := x2;
        yorig := y2;

        drawROIquad1;
        diam3a := halfdiam;
        selectwindow('Fibre diameters');
        writeln('Diam3a is ',diam3a);

        drawROIquad3;
        diam3b := halfdiam;
        selectwindow('Fibre diameters');
```

```
            writeln('Diam3b is ',diam3b);

        end;

end;

{————————————————————————————————————————}


macro 'AFM measure';

{This macro uses a user-drawn ROI that should be drawn along the long axis
    of a collagen fibre and measures its thickness and the banding
    periodicity. It does this by drawing secondary ROIs perpendicular to the
    first one and finds the first significant step in grey values along the
    secondary ROI. Adding the distances on both sides gives the presumed
    fibre diameter across one point of the user ROI. Repeating this along the
     ROI gives an average. The banding periodicity comes from an FFT of the
    user ROI and finding the major frequency in this one dimensional array.
    This last step has to be done by the user separately in the saved '
    filename-PV' file.}

var

picid1,picid2 : integer;
picwidth, picheight : integer;
type : integer;                              {ROI type}
title : string;
i : integer;
count, ppv, min, max, difflength : integer;
linelength, linelength1, linewidth : integer; {linelength is the extended
    power of 2 ROI, linelength1 is the length of the actual ROI drawn}
{expo, expochk : real;}

x1, y1, x2, y2 : integer;        {The start and end points of the original
    ROI}
x0, y0 : integer;              {The centre of the user's line ROI}

a, alpha : real;
b, beta : real;

quadr : integer;

diam1, diam2, diam3, avgdiam, stdevdiam : real;      {fibre diameters and
    stdev}


begin

        savestate;

        picid1 := pidnumber;

        doyinvert;

        type := get('ROItype');                  {ROI type 6 = straight line}

        if type <> 6 then begin
        putmessage('Please draw a line ROI along a collagen long axis. Press
    esc and draw a new ROI.');
        end;

        {picid1 := pidnumber;}
```

```
    getpicsize(picwidth, picheight);
    copy;                {copies current ROI onto clip board}

    choosepic(picid1);
    setscale(0,'pixel');
    title := windowtitle;
    delete(title, 15, 999);      {removes 999 characters from the title,
  starting at 15th}
    selectall;
    duplicate(title ' meas');
    picid2 := pidnumber;
    restoreROI;
    choosepic(picid1);
    killroi;
    choosepic(picid2);
    setoptions('x-y center, length, angle, rUser1');
    setprecision(2);
    resetcounter;

    measure;

    getline(x1,y1,x2,y2,linewidth);

    linelength := rLength[1];        {from Measure, will be set to power of
   2 later}
    linelength1 := rLength[1];
    a := rAngle[1];
    x0 := rX[1];              {this is X-Y Centre from Measure}
    y0 := rY[1];

    resetcounter;

    getplotdata(count,ppv,min,max);      {count is number of gvs!}

    newtextwindow('Fibre diameters',300,200);


{---------------- getting the period----------------}

    getperiod;

    choosepic(picid2);
    drawmainroi;

{------------------ getting the fibre thickness ----------------}

    choosepic(picid2);

    if (a >= 0) and (a < 90) then begin
          quadr := 1;
          alpha := a;
    end
    else
    if (a >= 90) and (a <180) then begin
          quadr := 2;
          alpha := 180 - a;
    end
    else
    if (a >= 180) and (a < 270) then begin
          quadr := 3;
          alpha := a - 180;
    end
    else
```

```
    if (a >= 270) and (a < 360) then begin
          quadr := 4;
          alpha := 360 - a;
    end;

    showmessage('Angle of ROI is: ',a);

    alpha := a * (3.14159265 / 180);      {this is to convert from degree
  into radians}
    b := abs(90 - a);
    beta := b * (3.14159265 / 180);

    drawallROIs;

    diam1 := diam1a + diam1b;
    diam2 := diam2a + diam2b;
    diam3 := diam3a + diam3b;

    avgdiam := (diam1 + diam2 + diam3)/3;
    stdevdiam := sqrt((((diam1 - avgdiam) * (diam1 - avgdiam)) + ((diam2 -
   avgdiam) * (diam2 - avgdiam)) + ((diam3 - avgdiam) * (diam3 - avgdiam)))
   / 2);

    selectwindow('Fibre diameters');
    writeln('Diameter 1 centre is ',diam1);
    writeln('Diameter 2 start is ',diam2);
    writeln('Diameter 3 end is ',diam3);
    writeln;
    writeln('Average diameter is ',avgdiam:5:0);
    writeln('Stdev is ',stdevdiam:3:2);

    restorestate;

end.
```

B.2 PLM Macro

```
macro 'PLM ACF macro';

{This macro is intended for polarised light microscopy images. It moves a
    user-defined ROI in steps across the image, calculates the ACF,
    thresholds this and copies and pastes the resulting particle into a new
    window. It also uses the particle analysis to identify minor and major
    axes and angles.}

var

picid1, picid2, picwidth, picheight, xo, yo, xe, ye : integer;
ROIw,ROIh : integer;
ROImidx, ROImidy, midx, midy : real;
m, n, i, j, l, lp, midp, loff, lpst : integer;
fft1, fft2 : integer;
a, b, c, delta : real;
title, datawindow, acfparticles : string;
thresh : integer;
fil : integer;
type : integer;
ROIbefore : integer   {this is introduced to implement Steve's suggestion to
     change the i and j loops, so that I don't lose the last column and row
    ...}
```

```
{-----------------------------------------------}

function power(a,b : integer;) : integer;
{calculates a to the power of b}

begin

  power := exp(b*ln(a));

end;


{-----------------------------------------------}

procedure ACFtilespart(ROIw : integer;);
{main part}


begin

  ROIh := ROIw;

  ROImidx := ROIw / 2.0;
  ROImidy := ROIh / 2.0;
  {These will be used to find the central particle and have to be real as
    the centre may be calculated as real.}

  SetNewSize(picwidth, picheight);

  delete(title,9,999);       {deletes the characters up to 999, starting at
    the 10th}

  acfparticles := concat(title,'ACF ROI',ROIw:4:0);

  MakeNewWindow(acfparticles);

  Picid2 := Pidnumber;

 m := trunc(Picwidth / ROIw);
 n := trunc(Picheight / ROIh);
 {it is necessary to use trunc, as div only works with integers, and what
   comes out of the power function is a real number!}

 {m and n are how many times the ROI goes into the image.}

 Resetcounter; {to make sure the measure array is reset to the start (this
   doesn't mean it's now empty)}


 l := 0;
 {l will be used to fill the rUser arrays.}

 lp := 1;
 {lp will be used to find the central particle}

 loff := 100;
 {This is an offset for the counter in the user arrays, as something else
   seems to use User5 and overwrite it.}

 for j := 0 to 2 * (n - 1) do begin

   for i := 0 to 2 * (m - 1) do begin
```

```
  l := l + 1;

  rUser1[l+loff] := -1;   {flags to show that no particle has been
detected, rUser1 is for AxialAnisotropy}
  rUser2[l+loff] := -1;   {rUser2 is for Angle}
  rUser3[l+loff] := 0;    {rUser3 is for i = x pos or column}
  rUser4[l+loff] := 0;    {rUser4 is for j = y pos or row}
  rUser5[l+loff] := 0;    {rUser5 is for area of minpoint particle}

  SetCounter(1);
  Setoptions('User1User2User3User4User5');

  Choosepic(picid1);

  MakeROI((i/2) * ROIw,(j/2) * ROIh,ROIw,ROIh);

  fft('foreward');
  fft1 := pidNumber;
  ImageMath('cmul', fft1, fft1, 1, 0, 'FFT2');
  fft2 := pidNumber;
  ChoosePic(fft1);
  Dispose;
  ChoosePic(fft2);
  fft('Inverse');
  fft('Swap Quadrants');
  SetPicName('Autocorrelation');

  Setthreshold(thresh);
  {threshold so far arbitrary - might find a way to relate it to the
histogram}

  ApplyLUT;

  SetOptions('X-Y Center Major Minor Angle');    {As long as XY is
ticked (default), it calculates the ellipse anyway!}

  for fil := 1 to 999 do begin
  rX[fil] := 0;
  rY[fil] := 0;
  end;
  {This is to empty the array.}


  AnalyzeParticles('reset');

  {The reset is there because the options in the brackets are optional,
but whatever isn't mentioned is disabled. So this is a way to switch off
the labelling of particles, which, when enabled, returns only 1s in the
new window instead of a particle.}

  SetThreshold(126);

  {Returns the particle to black. This is necessary because the
AutoOutline can only detect value 255 pixels!!}

    ApplyLUT;


      midp := -1;              {impossible option}
      for lp := 1 to rCount do begin
        {Putmessage('ROImidx=',ROImidx,'  ROImidy='ROImidy,'  rCount=',
rCount,'  lp=',lp,'   x=',rX[lp],'  y=',rY[lp]);}
  {The Putmessage is for diagnostic purposes.}
```

```
            if ((rX[lp] > (roimidx-2)) and (rX[lp] < (roimidx+2))) then
begin
              if((rY[lp] > (roimidy-2)) and (rY[lp] < (roimidy+2))) then
begin
                midp := lp;
              end;
            end;
        end;


        if(midp > 0) then begin
   {to make sure it draws nothing if there's no particle, without this
when there's no particle it would draw the result from the last particle
again}

     SelectWindow('Autocorrelation');

     midx := rX[midp];
     midy := rY[midp];

     Autooutline(midx,midy);

     {The macro version of the Wand tool. Can only detect black, not grey
!}


     type := get('ROItype');
     if type <> 0 then begin

       SaveROIArray(1);

       {We only need 1 because it gets overwritten the next time round
the loop.}

             Copy;

       {The SaveROIArray only saves the outline, not the content, so Copy
 copies the content of the ROI to the clipboard.}

       KillROI;


           Choosepic(picid2);

       RestoreROIArray(1);

       MoveROI(i/2 * ROIw,j/2 * ROIh);

             Paste;

             KillROI;

     end;        {end of if type <>0 ...}

     a := rMajor[midp];

         b := rMinor[midp];

         delta := rAngle[midp];

     rUser1[l+loff] := a / b;
     rUser2[l+loff] := delta;
     rUser3[l+loff] := i;
```

```
        rUser4[l+loff] := j;
        rUser5[l+loff] := rArea[midp];


      end;     {end of 'if (midp < 0) '...}

      SelectWindow('Autocorrelation');
      Dispose;


        KillROI;


    end;     {end of i loop}

  datawindow := concat(title,' data ROI',ROIw:4:0,j);

  NewTextWindow(datawindow,500,300);
  WriteLn('ACF Number  Ax. Anisotropy a/b  Angle   Col   Row  Area');
  WriteLn;

  lpst := l - (2 * (n-1));     {calculates the start of the new data file}

  for lp := lpst to l do begin
      writeln(' ', lp,' ', rUser1[lp+loff]:6:2,' ', rUser2[lp+loff]:3:1,' ',
    rUser3[lp+loff]:4:0,' ',rUser4[lp+loff]:4:0,' ',rUser5[lp+loff]:5:0);

  end;

  saveas('Macintosh HD:Users:Marion:folder:'Datawindow);
  close;




  end;     {end of j loop}

{ datawindow := concat(title,' data ROI',ROIw:4:0);

  NewTextWindow(datawindow,500,300);
  WriteLn('ACF Number  Ax. Anisotropy a/b  Angle   Col   Row  Area');
  WriteLn;

  for lp := 1 to l do begin
      writeln(' ', lp,' ', rUser1[lp+loff]:6:2,' ', rUser2[lp+loff]:3:1,' ',
    rUser3[lp+loff]:4:0,' ',rUser4[lp+loff]:4:0,' ',rUser5[lp+loff]:5:0);

  end;

  saveas('Macintosh HD:Users:Marion:folder:'Datawindow);
  close;}

  selectpic(picid2);

  saveas('Macintosh HD:Users:Marion:folder:'acfparticles);
  close;

  choosepic(picid1);
  killROI;
  {close;}


end;     {end that goes with initial begin}
```

```
{————————————————————————————————————}

procedure ACFtileswROIsizes(ROIstart, ROIend : integer;);

var

ROIsize, p : integer; {has to be a power of 2}


begin

  picid1 := PidNumber;

  SetScale(0,'pixel');
  {The image has to be in pixels for the ROImids to work.}

  title := Windowtitle;

  Getpicsize(picwidth, picheight);

  thresh := 150;

  {ROIw := Getnumber('Select ROI size (power of 2 square): ',128,0);}

  for p := ROIstart to ROIend do begin

    ROIsize := power(2,p);
    ROIbefore := power(2,p-1);
    ACFtilespart(ROIsize);

  end;


end;


{————————————————————————————————————}

Macro 'ACFtileswROIsizes all in folder';

var

fromROI, toROI : integer;
filenum : integer;
num : integer;
currimg : integer;
starttick, endtick, durationm, durationh : integer;


begin


  updatesoff;

  savestate;

  starttick := tickcount;

  fromROI := Getnumber('Select starting power of 2 (4 = ROI 16): ',4,0);
  toROI := Getnumber('Select final power of 2 (5 = ROI 32): ',5,0);
```

```
  filenum := Getnumber('Number of images: ',1,0);

  for num := 1 to filenum do begin

    currimg := pidnumber;
    selectPic(currimg);
    ACFtileswROIsizes(fromROI, toROI);

    selectPic(currimg);
    close;

  end;

  endtick := tickcount;
  durationh := trunc((endtick - starttick)/216000);
  durationm := trunc((endtick - starttick)/3600) - (durationh * 60);
  showmessage('This macro took ',durationh,' hours and ',durationm,' minutes
     to run.');

  restorestate;

  updateson;

end.
```

B.3 TEM Macro

```
macro 'TEM QEM RMnxn';

{This macro calculates the input numbers for the Quality Evaluation Number
    for a Rotating Mask (RM) -based analysis method for TEM images and
    displays the results.
The analysis consists of the application of a RM of size n (the diameter of
    the ROI in pixels, followed by erosion and dilation steps and a watershed
     segmentation.
Displayed are the number of blobs detected, the image area they cover and
    their average axial anisotropy.}

var

para : real;       {parameter to be examined (here ROI size)}
{has to be real for the user arrays}
i : integer;
picid1, picid2 : integer;
picwidth, picheight : integer;
blackpix : integer;                {black pixels counted}
areapercent : real;                {black pixels as percent of whole image}
wholearea : integer;                 {black and white pixels together}
mean : real;                     {mean gv of original}
nbin : integer;                    {counter for erosion/dilation iterations}
npart : real;                    {counter for all particles in image}
avgecc, ecc : real;                {average axial anisotropy of blobs}
words, winname : string;
startroi, endroi : integer;

begin

  startroi := 1;
  endroi := 15;
{These are the integers that make the odd roi sizes: 1 = 3 to 15 = 31. using
     these variables allows the printout at the end to just show the roi
    sizes used and not go from 1 to the biggest roi size.}
```

```
picid1 := Pidnumber;
selectpic(picid1);

getpicsize(picwidth, picheight);

{compensation('tilt');}

setscale(0,'pixel');}

resetcounter;

setoptions('mean');

measure;

mean := rMean[rCount];

resetcounter;        {this is so that all rCount within the loop start with
   the same value}

setprecision(2);

{setoptions('User1 user2  user3 user4'); woz ere}

setuser1label('ROIsize');
setuser2label('Numblobs');
setuser3label('Area%');
setuser4label('Avgecc');

for para := startroi to endroi do begin                              {loop
  for ROI sizes start}

  choosepic(picid1);

  Rotatingmask(2*para + 1);

  picid2 := Pidnumber;

  makeROI(32,32,picwidth-64,picheight-64);          {31 is largest ROI
  size I'll ever use in this, 64 = 2 * 32}

  duplicate(windowtitle);

  picid2 := pidnumber;                        {sets the smaller image as
  picid2}

  setthreshold(mean);

  applylut;

  smooth;
  smooth;
  smooth;

  setthreshold(mean);

  applylut;

  setbinarycount(3);

  for nbin := 1 to 10 do begin          {binary iteration}
```

```
    erode;

  end;

  for nbin := 1 to 10 do begin                {binary iteration}

    dilate;

  end;

  gethistogram(0,0,picwidth-64,picheight-64);

  blackpix := histogram[255];

  wholearea := histogram[0] + blackpix;              {this makes wholearea
  independent of image size}

  areapercent := blackpix * 100 / wholearea;

  rUser3[para] := areapercent;                                  {area
  percent occupied by blobs}
  choosepic(picid2);

  watershed;

  setoptions('major minor');

  setparticlesize(100,999999);                                {"(min, max)"}
  analyzeparticles('include');                {"ignore" for edge particles,
  "include" for internal holes}
  showresults;

  {close;}
  ecc := 0;

  for npart := 0 to rCount do begin              {start of loop through
  particle eccentricities}

    ecc := ecc + (rMajor[rCount] / rMinor[rCount]);

    avgecc := (ecc/rCount) * 100;                    {I'm multiplying by
  100 because I can't get ISXM to give me decimal points}

{this should calculate the average axial anisotropy as we go along}

    rUser4[para] := avgecc;
  {avg axial anisotropy of blobs}

  end;                            {end of loop through particle anisotropies
  }

  rUser2[para] := rCount;                                              {
  number of blobs}

  rUser1[para] := 2*para +1;                                          {ROI
  size}

  resetcounter;

end;
    {loop for ROI sizes end}

{setoptions('User1 user2 user3 user4');}
```

```
  setcounter(para);

  setprecision(2);

  {updateresults;}

  {showresults;}
 winname := 'OPSRMnxn.csv';
 NewTextWindow(winname,500,200);
 words := 'ROIsize, Blobnum, Area%, Avganis';
 writeln(words);

 for i := startroi to endroi do begin
    writeln(rUser1[i], ',', rUser2[i], ',', rUser3[i]:3:0, ',', rUser4[i
    ]:4:0);
 end;

 selectwindow(winname);
 SaveAs;

end.
```

References

[1] Olaf S. Andersen. "Introduction to Biophysics Week: What is Biophysics?" In: *Biophysical Journal* 110.5 (2016), E01–E03. ISSN: 00063495. DOI: 10.1016/j.bpj.2016.02.012.

[2] Robert O Becker and Gary Selden. *The Body Electric*. 1st Quill. William Morrow, 1985, p. 368. ISBN: 9780688069711.

[3] Herbert Fröhlich and Friedrich Kremer, eds. *Coherent Excitations in Biological Systems*. Berlin Heidelberg: Springer, 1983. ISBN: 9783642691874.

[4] Herbert Fröhlich, ed. *Biological Coherence and Response to External Stimuli*. Springer Science & Business Media, 1988. ISBN: 9783642733093.

[5] Gilbert Ling. *In Search of the Physical Basis of Life*. 1st. Springer US, 1984. ISBN: 978-1-4613-2667-0. DOI: 10.1007/978-1-4613-2667-0.

[6] Erwin Schrödinger. *What is Life?: With Mind and Matter and Autobiographical Sketches*. Cambridge University Press, 1992. ISBN: 9781139643047.

[7] A V Hill. "Why Biophysics?" In: *Science* 124.3234 (1956), pp. 1233–1237.

[8] Sönke Johnsen. *The Optics of Life : a Biologist's Guide to Light in Nature*. Princeton University Press, 2011. ISBN: 0691139911.

[9] Kelly Servick. "It will be much harder to call new findings 'significant' if this team gets its way". In: *Science* (July 2017). ISSN: 0036-8075. DOI: 10.1126/science.aan7154. URL: http://www.sciencemag.org/news/2017/07/it-will-be-much-harder-call-new-findings-significant-if-team-gets-its-way.

[10] V. S. Ramachandran and Sandra. Blakeslee. *Phantoms in the brain : human nature and the architecture of the mind*. Fourth Estate, 1999, p. 328. ISBN: 1857028953. URL: https://books.google.co.uk/books/about/Phantoms%7B%5C_%7Din%7B%5C_%7Dthe%7B%5C_%7DBrain.html?id=RdkfQAAACAAJ%7B%5C&%7Dredir%7B%5C_%7Desc=y.

[11] Kenneth N. Ogle. "On the Resolving Power of the Human Eye". In: *Journal of the Optical Society of America* 41.8 (Aug. 1951), p. 517. ISSN: 0030-3941. DOI: 10.1364/JOSA.41.000517. URL: https://www.osapublishing.org/abstract.cfm?URI=josa-41-8-517.

[12] Raphael Kopan et al. "disease A common enzyme connects Notch signaling and Alzheimer ' s disease". In: *Genes & Development* 46202.314 (2011), pp. 723–742. ISSN: 0363-6143. DOI: 10.1152/ajpcell.00462.2010..

[13] Andrew M.K. Law et al. “Andy’s Algorithms: New automated digital image analysis pipelines for FIJI”. In: *Scientific Reports* 7.1 (2017), pp. 1–11. ISSN: 20452322. DOI: 10.1038/s41598-017-15885-6.

[14] Fangfang Zhong et al. “A comparison of visual assessment and automated digital image analysis of Ki67 labeling index in breast cancer”. In: *PLoS ONE* 11.2 (2016), pp. 1–11. ISSN: 19326203. DOI: 10.1371/journal.pone.0150505.

[15] Alexis Ramfos, Andreas Gazis, and George Katselis. “Development and evaluation of an automated digital image analysis software for obtaining seagrass leaf metrics”. In: *Botanica Marina* 55.6 (2012), pp. 601–610. ISSN: 00068055. DOI: 10.1515/bot-2012-0132.

[16] Keith S Delaplane, Jozef Van Der Steen, and Ernesto Guzman-novoa. “Standard methods for estimating strength parameters of Apis mellifera colonies Métodos estándar para estimar parámetros sobre la fortaleza de las colonias de Apis mellifera”. In: *Journal of Apicultural Research* 52.1 (2013), pp. 1–12. ISSN: 00218839. DOI: 10.3896/IBRA.1.52.1.03.

[17] C. H. Bock et al. “Plant disease severity estimated visually, by digital photography and image analysis, and by hyperspectral imaging”. In: *Critical Reviews in Plant Sciences* 29.2 (2010), pp. 59–107. ISSN: 07352689. DOI: 10.1080/07352681003617285.

[18] Martin Stevens, Alice E. Lown, and Alexander M. Denton. “Rockpool gobies change colour for camouflage”. In: *PLoS ONE* 9.10 (2014). ISSN: 19326203. DOI: 10.1371/journal.pone.0110325.

[19] I. Hunt von Herbing and S. M. Gallager. “Foraging behavior in early Atlantic cod larvae (Gadus morhua) feeding on a protozoan (Balanion sp.) and a copepod nauplius (Pseudodiaptomus sp.)” In: *Marine Biology* 136.3 (2000), pp. 591–602. ISSN: 0025-3162. DOI: 10.1007/s002270050719. URL: http://link.springer.com/10.1007/s002270050719.

[20] Shayesteh Sepehr et al. “Unsaturated fatty acid, cis-2-decenoic acid, in combination with disinfectants or antibiotics removes pre-established biofilms formed by food-related bacteria”. In: *PLoS ONE* 9.7 (2014). ISSN: 19326203. DOI: 10.1371/journal.pone.0101677.

[21] Torsten Böker et al. “A Hubble Space Telescope Census of Nuclear Star Clusters in Late-Type Spiral Galaxies. I. Observations and Image Analysis”. In: *The Astronomical Journal* 123.3 (2002), pp. 1389–1410. ISSN: 00046256. DOI: 10.1086/339025. arXiv: 0112086v1 [arXiv:astro-ph]. URL: http://adsabs.harvard.edu/abs/2002AJ....123.1389B.

[22] David Landgrebe. “Hyperspectral image data analysis”. In: *IEEE Signal Processing Magazine* 19.1 (2002), pp. 17–28. ISSN: 10535888. DOI: 10.1109/79.974718.

[23] Sefer B. Coskun and Norman C. Wardlaw. "Image analysis for estimating ultimate oil recovery efficiency by waterflooding for two sandstone reservoirs". In: *Journal of Petroleum Science and Engineering* 15.2-4 (1996), pp. 237–250. ISSN: 0920-4105. DOI: 10.1016/0920-4105(96)00002-2. URL: http://linkinghub.elsevier.com/retrieve/pii/0920410596000022.

[24] M C Cooper. "The use of digital image analysis in the study of laminated sediments". In: (1997), pp. 33–40.

[25] E Bradley, D Roberts, and C Still. "Environmental Modelling & Software Design of an image analysis website for phenological and meteorological monitoring". In: *Environmental Modelling and Software* 25.1 (2010), pp. 107–116. ISSN: 1364-8152. DOI: 10.1016/j.envsoft.2009.07.006. URL: http://dx.doi.org/10.1016/j.envsoft.2009.07.006.

[26] Alfredo D. Collado, Emilio Chuvieco, and Ana Camarasa. "Satellite remote sensing analysis to monitor desertification processes in the crop-rangeland boundary of Argentina". In: *Journal of Arid Environments* 52.1 (2002), pp. 121–133. ISSN: 01401963. DOI: 10.1006/jare.2001.0980. URL: http://linkinghub.elsevier.com/retrieve/pii/S0140196301909802.

[27] Zhaohua Chen et al. "Digital Photograph Analysis for Measuring Percent Plant Cover in the Arctic". In: 63.3 (2010), pp. 315–326.

[28] Yushin Ahn, Bumshick Shin, and Kyu-Han Kim. "Shoreline Change Monitoring using High Resolution Digital Photogrammetric Technique". In: *Journal of Coastal Research* 79 (2017), pp. 204–208. ISSN: 0749-0208. DOI: 10.2112/SI79-042.1. URL: http://www.bioone.org/doi/10.2112/SI79-042.1.

[29] Cinzia Perrino and Francesca Marcovecchio. "A new method for assessing the contribution of Primary Biological Atmospheric Particles to the mass concentration of the atmospheric aerosol". In: *Environment International* 87 (2016), pp. 108–115. ISSN: 18736750. DOI: 10.1016/j.envint.2015.11.015. URL: http://dx.doi.org/10.1016/j.envint.2015.11.015.

[30] Rachael M. Pentney and Mark E. Dickson. "Digital Grain Size Analysis of a Mixed Sand and Gravel Beach". In: *Journal of Coastal Research* 279.1 (2012), pp. 196–201. ISSN: 0749-0208. DOI: 10.2112/JCOASTRES-D-10-00046.1. URL: http://www.bioone.org/doi/abs/10.2112/JCOASTRES-D-10-00046.1.

[31] Nigussie Haregeweyn et al. "Dynamics of land use and land cover and its effects on hydrologic responses: case study of the Gilgel Tekeze catchment in the highlands of Northern Ethiopia". In: *Environmental Monitoring and Assessment* 187.1 (2015). ISSN: 15732959. DOI: 10.1007/s10661-014-4090-1.

[32] Saied Pirasteh, Hojjat O Safari, and Somayeh Mollaee. "Monitoring and Modeling of Global Changes: A Geomatics Perspective". In: (2015), pp. 281–299. DOI: 10.1007/978-94-017-9813-6. URL: http://link.springer.com/10.1007/978-94-017-9813-6.

[33] Timo Ahonen, Abdenour Hadid, and Matti Pietikäinen. "Face description with local binary patterns: Application to face recognition". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28.12 (2006), pp. 2037–2041. ISSN: 01628828. DOI: 10.1109/TPAMI.2006.244.

[34] Aleš Muroň, Petr Koiš, and Jaroslav Pospíšil. "Identification of persons by means of the Fourier spectra of the optical transmission binary models of the human irises". In: *Optics Communications* 192.3-6 (2001), pp. 161–167. ISSN: 00304018. DOI: 10.1016/S0030-4018(01)01169-5.

[35] Wojciech Sankowski et al. "Reliable algorithm for iris segmentation in eye image". In: *Image and Vision Computing* 28.2 (2010), pp. 231–237. ISSN: 02628856. DOI: 10.1016/j.imavis.2009.05.014. URL: http://dx.doi.org/10.1016/j.imavis.2009.05.014.

[36] B. M. Mehtre. "Fingerprint image analysis for automatic identification". In: *Machine Vision and Applications* 6.2-3 (1993), pp. 124–139. ISSN: 09328092. DOI: 10.1007/BF01211936.

[37] Miguel Ángel Rogerio-Candelera. "Digital image analysis based study, recording, and protection of painted rock art. Some Iberian experiences". In: *Digital Applications in Archaeology and Cultural Heritage* 2.2-3 (2015), pp. 68–78. ISSN: 22120548. DOI: 10.1016/j.daach.2014.11.001. URL: http://dx.doi.org/10.1016/j.daach.2014.11.001.

[38] Martin D Brasier et al. "Questioning the evidence for Earth ' s oldest fossils". In: *Nature* 416.March (2002), pp. 76–81.

[39] Dr David Compton Dr Simon Bramble and Ms Lena Klasén. "Forensic Image Analysis version 2". In: *Forensic, Interpol Symposium, Science* (2001).

[40] Marie Spohn, Andrea Carminati, and Yakov Kuzyakov. "Soil zymography - A novel in situ method for mapping distribution of enzyme activity in soil". In: *Soil Biology and Biochemistry* 58 (2013), pp. 275–280. ISSN: 00380717. DOI: 10.1016/j.soilbio.2012.12.004. URL: http://dx.doi.org/10.1016/j.soilbio.2012.12.004.

[41] J. E. Lundquist and R. A. Sommerfeld. "Use of fourier transforms to define landscape scales of analysis for disturbances: A case study of thinned and unthinned forest stands". In: *Landscape Ecology* 17.5 (2002), pp. 445–454. ISSN: 09212973. DOI: 10.1023/A:1021252222319.

[42] H P Schwarz et al. "Development of an autonomous driven robotic platform used for ht-phenotyping in viticulture". In: *Mech.Eng.Lett.Szent István Univ* 10.November (2013), pp. 153–160.

[43] Roland T Chin and A N I Charles A Harlow. "Automated Visual Inspection : A". In: 6 (1982), pp. 557–573.

[44] Douglas F. Barbin et al. “Digital image analyses as an alternative tool for chicken quality assessment”. In: *Biosystems Engineering* 144 (2016), pp. 85–93. ISSN: 15375110. DOI: 10.1016/j.biosystemseng.2016.01.015. URL: http://dx.doi.org/10.1016/j.biosystemseng.2016.01.015.

[45] A Bastawros and R McManuis. “Case Study: Use of Digital Image Analysis Software to Measure Non-uniform Deformation in Cellular Aluminium Alloys.” In: 22.2 (1998), pp. 35–37.

[46] J. Paulo Davim, J. Campos Rubio, and A. M. Abrao. “A novel approach based on digital image analysis to evaluate the delamination factor after drilling composite laminates”. In: *Composites Science and Technology* 67.9 (2007), pp. 1939–1945. ISSN: 02663538. DOI: 10.1016/j.compscitech.2006.10.009.

[47] S. Anbazhagan, N. K. Sainaba, and S. Arivazhagan. “Remote Sensing Study of Sittampundi Anorthosite Complex, India”. In: *Journal of the Indian Society of Remote Sensing* 40.1 (2012), pp. 145–153. ISSN: 0255660X. DOI: 10.1007/s12524-011-0126-y.

[48] R F Yu, W P Cheng, and M-L Chu. “On-line Monitoring of Wastewater True Color Using Digital Image Analysis and Artificial Neural Network”. In: *Journal of Environmental Engineering* ?Jan (2005), pp. 71–80. ISSN: 07339372. DOI: 10.1061/(ASCE)0733-9372(2005)131.

[49] Brainden. *Color Illusions*. URL: http://brainden.com/color-illusions.htm%7B%5C#%7DprettyPhoto[pp%7B%5C_%7Dgal]/4/ (visited on 10/20/2017).

[50] Steve Barrett. *Image SXM*. URL: https://www.imagesxm.org.uk (visited on 11/25/2017).

[51] Wayne S Rasband. *NIH Image Home Page*. URL: https://imagej.nih.gov/nih-image/ (visited on 11/25/2017).

[52] Vincent R Sherman, Wen Yang, and Marc A Meyers. “The materials science of collagen”. In: *Journal of the Mechanical Behavior of Biomedical Materials* 52 (2015), pp. 22–50. ISSN: 1751-6161. DOI: 10.1016/j.jmbbm.2015.05.023.

[53] Peter Fratzl, ed. *Collagen: Structure and Mechanics*. Boston, MA: Springer Science & Business Media, 2008. ISBN: 978-0-387-73905-2. DOI: 10.1007/978-0-387-73906-9.

[54] Johanna Myllyharju and Kari I Kivirikko. “Collagens , modifying enzymes and their mutations in humans , flies and worms”. In: *Trends in Genetics* 20.1 (2004), pp. 33–43. DOI: 10.1016/j.tig.2003.11.004.

[55] Johanna Myllyharju et al. “Collagens and collagen-related diseases”. In: *Annals of Medicine* 33.1 (2001), pp. 7–21. DOI: 10.3109/07853890109002055.

[56] Karl E Kadler et al. “Collagens at a Glance”. In: *Journal of Cell Science* 120.12 (2007), pp. 1955–1958. DOI: 10.1242/jcs.03453.

[57] Sylvie Ricard-Blum and Florence Ruggiero. “The collagen superfamily: From the extracellular matrix to the cell membrane”. In: *Pathologie Biologie* 53.7 (2005), pp. 430–442. DOI: 10.1016/j.patbio.2004.12.024.

[58] Michael J Mienaltowski and David E Birk. “Structure, Physiology and Biochemistry of Collagens”. In: *Progress in Heritable Soft Connective Tissue Diseases, Advances in Experimental Medicine and Biology*. Ed. by J Halper. January. Springer Science & Business Media, 2014. ISBN: 9789400778931. DOI: 10.1007/978-94-007-7893-1.

[59] Elizabeth G Canty et al. “Coalignment of Plasma Membrane Channels and Protrusions (Fibripositors) Specifies the Parallelism of Tendon”. In: *The Journal of Cell Biology* 165.4 (2004), pp. 553–563. DOI: 10.1083/jcb.200312071.

[60] D F Holmes et al. “STEM / TEM studies of collagen fibril assembly”. In: *Micron* 32 (2001), pp. 273–285.

[61] J W Smith. “Molecular Pattern in Native Collagen”. In: *Nature* 219.July 13 (1968), pp. 157–158.

[62] D A D Parry and A S Craig. “Electron microscope evidence for an 80 A unit in collagen fibrils”. In: *Nature* 282.8 Nov (1979), pp. 213–215.

[63] Joseph P R O Orgel et al. “The in situ Supermolecular Structure of Type I Collagen”. In: *Structure* 9.01 (2001), pp. 1061–1069.

[64] A S Craig et al. “An estimate of the mean length of collagen fibrils in rat tail tendon as a function of age”. In: *Connective Tissue Research* 19.1 (1989), pp. 51–62.

[65] Paolo P Provenzano and Ray Vanderby Jr. “Collagen fibril morphology and organization : Implications for force transmission in ligament and tendon”. In: *Matrix Biology* 25 (2006), pp. 71–84. DOI: 10.1016/j.matbio.2005.09.005.

[66] Rene B Svensson et al. “Evidence of structurally continuous collagen fibrils in tendons”. In: *Acta Biomaterialia* 50 (2017), pp. 293–301. DOI: 10.1016/j.actbio.2017.01.006.

[67] David E Birk and Robert L Trelstad. “Extracellular Compartments in Tendon Morphogenesis :” in: *The Journal of Cell Biology* 103.July (1986), pp. 231–240.

[68] Sameer Varma, Joseph P R O Orgel, and Jay D Schieber. “Nanomechanics of Type I Collagen”. In: *Biophysical Journal* 111.1 (2016), pp. 50–56. ISSN: 0006-3495. DOI: 10.1016/j.bpj.2016.05.038.

[69] Thomas Gutsmann et al. “Evidence that collagen fibrils in tendons are inhomogeneously structured in a tubelike manner.” In: *Biophysical Journal* 84.4 (2003), pp. 2593–2598. DOI: 10.1016/S0006-3495(03)75064-4.

[70] Laurent Bozec, Gert van der Heijden, and Michael Horton. “Collagen fibrils: nanoscale ropes.” In: *Biophysical Journal* 92.1 (2007), pp. 70–75. DOI: 10.1529/biophysj.106.085704.

[71] E G Canty and K E Kadler. “Collagen fibril biosynthesis in tendon : a review and recent insights”. In: *Comparative Biochemistry and Physiology A* 133 (2002), pp. 979–985.

[72] John E Scott. “Proteoglycan-fibrillar collagen Interactions”. In: *Biochemical Journal* 252 (1988), pp. 313–323.

[73] P Kannus. "Structure of the tendon connective tissue". In: *Scandinavian Journal of Medical Science Sports* 10.3 (2000), pp. 312–320.

[74] Baptiste Depalle et al. "Influence of cross-link structure, density and mechanical properties in the mesoscale deformation mechanisms of collagen fibrils". In: *Journal of the Mechanical Behavior of Biomedical Materials* 52 (2015), pp. 1–13. DOI: 10.1016/j.jmbbm.2014.07.008.

[75] J Guimberteau. "The Microvacuolar System: How Connective Tissue Sliding Works". In: *The Journal of Hand Surgery / European Volume* 35E.8 (2010), pp. 614–622.

[76] J Guimberteau and Colin Armstrong. *The Architecture of Living Fascia: The Extracellular Matrix and Cells Revealed Through Endoscopy*. Handspring Publishing, 2015, p. 232. ISBN: 978-1909141117.

[77] J Guimberteau. *Strolling under the Skin - YouTube*. URL: https://www.youtube.com/watch?v=eW0lvOVKDxE%7B%5C&%7Dfeature=youtu.be (visited on 11/22/2017).

[78] *File:Diagram of human eye without labels.svg - Wikimedia Commons*. URL: https://commons.wikimedia.org/wiki/File:Diagram%7B%5C_%7Dof%7B%5C_%7Dhuman%7B%5C_%7Deye%7B%5C_%7Dwithout%7B%5C_%7Dlabels.svg (visited on 11/22/2017).

[79] Peter G. Watson and Robert D. Young. "Scleral structure, organisation and disease. A review". In: *Experimental Eye Research* 78.3 (2004), pp. 609–623. DOI: 10.1016/S0014-4835(03)00212-4.

[80] Y. Komai and T. Ushiki. "The three-dimensional organization of collagen fibrils in the human cornea and sclera". In: *Investigative Ophthalmology and Visual Science* 32.8 (1991), pp. 2244–2258. ISSN: 01460404.

[81] K. M. Meek. "The cornea and sclera". In: *Collagen: Structure and Mechanics* (2008), pp. 359–396. ISSN: 0020-8167. DOI: 10.1007/978-0-387-73906-9_13.

[82] G E Marshall, A G P Konstas, and Williarn R Lee. "Collagen in the aged human macular sclera". In: *Current Eye Research* 12.2 (1993), pp. 143–153.

[83] Neville a. McBrien and Alex Gentle. "Role of the sclera in the development and pathological complications of myopia". In: *Progress in Retinal and Eye Research* 22.3 (2003), pp. 307–338. ISSN: 13509462. DOI: 10.1016/S1350-9462(02)00063-0.

[84] Gerd Binnig and Heinrich Rohrer. "Scanning Tunneling Microscopy—from Birth to Adolescence (Nobel Lecture)". In: *Angewandte Chemie International Edition in English* 26.7 (1987), pp. 606–614. ISSN: 15213773. DOI: 10.1002/anie.198706061. arXiv: 0308464 [cond-mat].

[85] Franz J. Giessibl. "Advances in atomic force microscopy". In: *Reviews of Modern Physics* 75.3 (2003), pp. 949–983. ISSN: 00346861. DOI: 10.1103/RevModPhys.75.949. arXiv: 0305119 [cond-mat].

[86] G. Binnig and C. F. Quate. “Atomic Force Microscope”. In: *Physical Review Letters* 56.9 (1986), pp. 930–933. ISSN: 0031-9007. DOI: 10.1103/PhysRevLett.56.930.

[87] Q. Zhong et al. “Fractured polymer/silica fiber surface studied by tapping mode atomic force microscopy”. In: *Surface Science* 290 (1993), pp. L688–L692. ISSN: 00396028. DOI: 10.1016/0039-6028(93)90582-5. arXiv: arXiv:1303.3229v1.

[88] Y. Martin, C. C. Williams, and H. K. Wickramasinghe. “Atomic force microscope-force mapping and profiling on a sub 100-Å scale”. In: *Journal of Applied Physics* 61.10 (1987), pp. 4723–4729. ISSN: 00218979. DOI: 10.1063/1.338807.

[89] T. R. Albrecht et al. “Frequency modulation detection using high-Q cantilevers for enhanced force microscope sensitivity”. In: *Journal of Applied Physics* 69.2 (1991), pp. 668–673. ISSN: 00218979. DOI: 10.1063/1.347347. arXiv: arXiv:1011.1669v3.

[90] Mechanical Property Mapping. *Application Note # 128 Quantitative Mechanical Property Mapping at the Nanoscale with PeakForce QNM*. (Visited on 11/25/2017).

[91] Florian Rehfeldt and Christoph F Schmidt. “Physical probing of cells”. In: *Journal of Physics D: Applied Physics* 50.46 (2017), p. 463001. ISSN: 0022-3727. DOI: 10.1088/1361-6463/aa8aa6.

[92] Angelica R Harper and Jody A Summers. *The Dynamic Sclera: Extracellular matrix remodeling in normal ocular growth and myopia development*. 2015. DOI: 10.1016/j.exer.2014.07.015.

[93] Jouni Uitto and Jack R Lichtenstein. “Defects in the Biochemistry of Collagen in Diseases of Connective Tissue”. In: *Journal of Investigative Dermatology* 66.2 (1976), pp. 59–79. DOI: 10.1111/1523-1747.ep12481404.

[94] Mary Brewer and Thomas Scott. *Concise Encyclodedia of Biochemistry*. Berlin: Walter de Gruyter, 1983. ISBN: 3110078600.

[95] Margaret Tzaphlidou. “The role of collagen and elastin in aged skin: An image processing approach”. In: *Micron* 35.3 (2004), pp. 173–177. ISSN: 09684328. DOI: 10.1016/j.micron.2003.11.003.

[96] Margaret Tzaphlidou. “The role of collagen in bone structure : An image processing approach”. In: *Micron* 36 (2005), pp. 593–601. DOI: 10.1016/j.micron.2005.05.009.

[97] J. Bella. “Collagen structure: new tricks from a very old dog”. In: *Biochemical Journal* 473.8 (2016), pp. 1001–1025. ISSN: 0264-6021. DOI: 10.1042/BJ20151169.

[98] L C Junqueira, G S Montes, and E M Sanchez. “The influence of tissue section thickness on the study of collagen by the Picrosirius-polarization method.” In: *Histochemistry* 74.1 (1982), pp. 153–156. ISSN: 0301-5564. DOI: 10.1007/BF00495061.

[99] I Sobel and G Feldman. “A 3x3 isotropic gradient operator for image processing.” In: *in Hart, P. E. & Duda R. O. Pattern Classification and Scene Analysis* (1973), pp. 271–272. URL: papers2://publication/uuid/F6C98D8E-0A99-40EF-A91C-0ECA53448D1F.

[100] Keith M Meek and David F Holmes. "Interpretation of the electron microscopical appearance of collagen fibrils from corneal stroma". In: *International Journal of Biological Macromolecules* 5 (1983), pp. 17–25.

[101] *File:Light-wave.svg - Wikimedia Commons*. URL: https://commons.wikimedia.org/wiki/File:Light-wave.svg (visited on 11/26/2017).

[102] *File:Circular.Polarization.Circularly.Polarized.Light.With.Components.Right.Handed.svg - Wikimedia Commons*. URL: https://commons.wikimedia.org/wiki/File:Circular.Polarization.Circularly.Polarized.Light%7B%5C_%7DWith.Components%7B%5C_%7DRight.Handed.svg (visited on 11/26/2017).

[103] M M Swann and J M Mitchison. "Refinements in polarized light microscopy." In: *The Journal of Experimental Biology* 27.2 (1950), pp. 226–37. ISSN: 0022-0949.

[104] *ZEISS Microscopy Online Campus — Microscopy Basics — Enhancing Contrast in Transmitted Light*. URL: http://zeiss-campus.magnet.fsu.edu/articles/basics/contrast.html (visited on 11/26/2017).

[105] M. A. Geday et al. "Images of absolute retardance L·Δn, using the rotating polariser method". In: *Journal of Microscopy* 198.1 (2000), pp. 1–9. ISSN: 00222720. DOI: 10.1046/j.1365-2818.2000.00687.x.

[106] A Changoor et al. "Structural characteristics of the collagen network in human normal , degraded and repair articular cartilages observed in polarized light and scanning electron microscopies". In: *Osteoarthritis and Cartilage* 19.12 (2011), pp. 1458–1468. DOI: 10.1016/j.joca.2011.09.007.

[107] S. Kinoshita, S. Yoshioka, and J. Miyazaki. "Physics of structural colors". In: *Reports on Progress in Physics* (2008). ISSN: 00344885. DOI: 10.1088/0034-4885/71/7/076401.

[108] Andrew Richard Parker et al. "515 Million Years of Structural Colour". In: *Journal of Optics A: Pure Applied Optics* 2 (2000), R15–R28.

[109] Serge Berthier, Magali Thomé, and Priscilla Simonis. "Circular polarization in nature: Factual, theoretical and experimental summary". In: *Materials Today: Proceedings* 1S (2014), pp. 145–154. ISSN: 22147853. DOI: 10.1016/j.matpr.2014.09.015.

[110] L. C U Junqueira, G. Bignolas, and R. R. Brentani. "Picrosirius staining plus polarization microscopy, a specific method for collagen detection in tissue sections". In: *The Histochemical Journal* 11.4 (1979), pp. 447–455. ISSN: 00182214. DOI: 10.1007/BF01002772.

[111] L F Nielsen et al. "Sirius red and acid fuchsin staining mechanisms." In: *Biotechnic & histochemistry : official publication of the Biological Stain Commission* 73.2 (Mar. 1998), pp. 71–77. ISSN: 1052-0295. URL: http://www.ncbi.nlm.nih.gov/pubmed/9605621.

[112] John A Kiernan. *Sirius Red for Collagen Staining Protocol*. URL: http://medicine.emory.edu/documents/mimcore/sirius-red-collagen-staining-protocol.pdf (visited on 09/03/2017).

[113] D. Dayan et al. “Are the polarization colors of Picrosirius red-stained collagen determined only by the diameter of the fibers?” In: *Histochemistry* (1989). ISSN: 03015564. DOI: 10.1007/BF00266843.

[114] Renée Heilbronner and Steve Barrett. *Image Analysis in Earth Sciences: Microstructures and Textures of Earth Materials*. Springer Science & Business Media, 2013. ISBN: 364210343X.

[115] William H Press et al. *Numerical Recipes: The art of Scientific Computing*. 1st. Cambridge University Press, 1986, p. 848. ISBN: 978-0521308113.

[116] Mihaela Balu et al. “Effect of excitation wavelength on penetration depth in nonlinear optical microscopy of turbid media”. In: *Journal of Biomedical Optics* 14.1 (2009), p. 010508. ISSN: 10833668. DOI: 10.1117/1.3081544. arXiv: NIHMS150003. URL: http://biomedicaloptics.spiedigitallibrary.org/article.aspx?doi=10.1117/1.3081544.

[117] Paul J. Campagnola et al. “Three-dimensional high-resolution second-harmonic generation imaging of endogenous structural proteins in biological tissues”. In: *Biophysical Journal* 82.1 (2002), pp. 493–508. ISSN: 00063495. DOI: 10.1016/S0006-3495(02)75414-3. URL: http://dx.doi.org/10.1016/S0006-3495(02)75414-3.

[118] Justin C. Williams and Paul J. Campagnola. “Wearable Second Harmonic Generation Imaging: The Sarcomeric Bridge to the Clinic”. In: *Neuron* 88.6 (2015), pp. 1067–1069. ISSN: 10974199. DOI: 10.1016/j.neuron.2015.12.009. URL: http://dx.doi.org/10.1016/j.neuron.2015.12.009.

[119] P. Fratzl, N. Fratzl-Zelman, and K. Klaushofer. “Collagen packing and mineralization. An x-ray scattering investigation of turkey leg tendon”. In: *Biophysical Journal* 64.1 (1993), pp. 260–266. ISSN: 00063495. DOI: 10.1016/S0006-3495(93)81362-6. URL: http://dx.doi.org/10.1016/S0006-3495(93)81362-6.

[120] G. J. Cameron et al. “Structure of type I and type III heterotypic collagen fibrils: An X-ray diffraction study”. In: *Journal of Structural Biology* 137.1-2 (2002), pp. 15–22. ISSN: 10478477. DOI: 10.1006/jsbi.2002.4459.

[121] T J Wess et al. “A Consensus for Molecular packing of type I collagen in tendon.” In: *Journal of molecular biology* 275.2 (1998), pp. 255–67. ISSN: 0022-2836. DOI: 10.1006/jmbi.1997.1449. URL: http://www.sciencedirect.com/science/article/pii/S0022283697914492.

[122] Maxime Rivard et al. “The structural origin of second harmonic generation in fascia”. In: *Biomedical Optics Express* 2.1 (2011), p. 26. ISSN: 2156-7085. DOI: 10.1364/BOE.2.000026. URL: https://www.osapublishing.org/boe/abstract.cfm?uri=boe-2-1-26.

[123] Israel Rocha-Mendoza et al. “Sum frequency vibrational spectroscopy: The molecular origins of the optical second-order nonlinearity of collagen”. In: *Biophysical Journal* 93.12 (2007), pp. 4433–4444. ISSN: 00063495. DOI: 10.1529/biophysj.107.111047.

[124] A Deniset-Besseau et al. “Nonlinear optical response of the collagen triple helix and second harmonic microscopy of collagen liquid crystals”. In: *Multiphoton Microscopy in the Biomedical Sciences X* 7569 (2010), 75691S–5. ISSN: 16057422. DOI: Artn75691s\rDoi10.1117/12.840873. URL: http://link.aip.org/link/?PSI/7569/75691S/1.

[125] Xiyi Chen et al. “Second harmonic generation microscopy for quantitative analysis of collagen fibrillar structure”. In: *Nature Protocols* 7.4 (2012), pp. 654–669. ISSN: 1754-2189. DOI: 10.1038/nprot.2012.009. arXiv: NIHMS150003.

[126] Rebecca Williams. *Application Note #2 Second Harmonic Generation (SHG) Microscopy: The Forward-to- Backward (F/B) issue*. URL: http://www.drbio.cornell.edu/ApplicationNotes/SHG%7B%5C_%7DImaging%7B%5C_%7DFvsB.pdf (visited on 09/18/2017).

[127] Warren R. Zipfel, Rebecca M. Williams, and Watt W. Webb. *Nonlinear magic: Multiphoton microscopy in the biosciences*. 2003. DOI: 10.1038/nbt899.

[128] Vladimir Hovhannisyan et al. “Spatial orientation mapping of fibers using polarization-sensitive second harmonic generation microscopy.” In: *Journal of Biophotonics* 5.10 (2012), pp. 768–76. ISSN: 1864-0648. DOI: 10.1002/jbio.201100123.

[129] Paul Campagnola. “Second harmonic generation imaging microscopy: Applications to diseases diagnostics”. In: *Analytical Chemistry* 83.9 (2011), pp. 3224–3231. ISSN: 00032700. DOI: 10.1021/ac1032325.

[130] P. A. Franken et al. “Generation of optical harmonics”. In: *Physical Review Letters* 7.4 (1961), pp. 118–119. ISSN: 00319007. DOI: 10.1103/PhysRevLett.7.118.

[131] R Hellwarth and P Christensen. “Nonlinear Optical Microscope using Second-Harmonic Generation”. In: *Applied Optics* 14.2 (1975), pp. 247–248. DOI: 10.1364/AO.14.000247.

[132] J. N. Gannaway and C. J R Sheppard. “Second-harmonic imaging in the scanning optical microscope”. In: *Optical and Quantum Electronics* 10.5 (1978), pp. 435–439. ISSN: 03068919. DOI: 10.1007/BF00620308.

[133] I Freund, M Deutsch, and A Sprecher. “Connective tissue polarity. Optical second-harmonic microscopy, crossed-beam summation, and small-angle scattering in rat-tail tendon.” In: *Biophysical Journal* 50.4 (1986), pp. 693–712. ISSN: 00063495. DOI: 10.1016/S0006-3495(86)83510-X.

[134] Guy Cox. “Biological applications of second harmonic imaging”. In: *Biophysical Reviews* 3.3 (2011), pp. 131–141. ISSN: 18672450. DOI: 10.1007/s12551-011-0052-9.

[135] Raphaël Turcotte et al. “Molecular Order of Arterial Collagen Using Circular Polarization Second-Harmonic Generation Imaging”. In: *Biophysical Journal* 110.3 (2016), pp. 530–533. ISSN: 15420086. DOI: 10.1016/j.bpj.2015.12.030.

[136] Ulf G Indahl and Tormod Naes. “Evaluation of Alternative Spectral Feature Extraction Methods of Textural Images for Multivariate Modelling”. In: *Journal of Chemometrics* 12 (1998), pp. 261–279.

[137] Leila Mostaço-Guidolin, Nicole L. Rosin, and Tillie Louise Hackett. "Imaging collagen in scar tissue: Developments in second harmonic generation microscopy for biomedical applications". In: *International Journal of Molecular Sciences* 18.1772 (2017). ISSN: 14220067. DOI: 10.3390/ijms18081772.

[138] *Limits to Resolution in the Electron Microscope*. URL: http://www.ou.edu/research/electron/bmz5364/resolutn.html (visited on 11/26/2017).

[139] *File:Scheme TEM en.png - Wikimedia Commons*. URL: https://commons.wikimedia.org/wiki/File:Scheme%7B%5C_%7DTEM%7B%5C_%7Den.png (visited on 11/26/2017).

[140] Volker Springel et al. "Simulations of the formation, evolution and clustering of galaxies and quasars". In: *Nature* 435.7042 (2005), pp. 629–636. ISSN: 0028-0836. DOI: 10.1038/nature03597. arXiv: 0504097 [astro-ph]. URL: http://www.nature.com/doifinder/10.1038/nature03597.

[141] J Roerdink and A Meijster. "The Watershed Transform: Definitions, Algorithms and Parallelization Strategies". In: *Fundamenta Informaticae* 41.1-2 (2001), pp. 187–228. ISSN: 0169-2968. DOI: 10.3233/FI-2000-411207. arXiv: arXiv:1011.1669v3.

[142] Richard O. Duda and Peter E. Hart. "Use of the Hough transformation to detect lines and curves in pictures". In: *Communications of the ACM* 15.1 (1972), pp. 11–15. ISSN: 00010782. DOI: 10.1145/361237.361242. URL: http://portal.acm.org/citation.cfm?doid=361237.361242.

[143] D. H. Ballard. "Generalizing the Hough transform to detect arbitrary shapes". In: *Pattern Recognition* 13.2 (1981), pp. 111–122. ISSN: 00313203. DOI: 10.1016/0031-3203(81)90009-1.